

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 8, 2003, 21:46:58 ; Search time 209 Seconds

(without alignments)  
872.784 Million cell updates/sec

Title: US-09-599-087b-5

Perfect score: 442

Sequence: 1 MRLVLSLLCILLCFSTF.....PKLEPEPLMVPALPQV 81

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 710428

Minimum DB seq length: 25  
Maximum DB seq length: 80

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL-frame+ p2n.model -DEV-xip  
-O/cq2.1/USPTO/US09599087/rnaut\_07052003.113320.7333/app\_query.fasta\_1.263  
-DB=N.Geneseq.101002 -QFMT=fastap -SUFFIX=mg -MINMATCH=0.1 -LOOPC=0  
-LOOPEXT=0 -UTITS=bits -START=1 -END=1 -MATRIX=blonum62 -TRANS=humanr0.cdi  
-LIST=45 -DOCLIN=200 -THR\_SCORE=pcct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=25 -MAXLEN=80  
-USER=US09599087.ecgn.1.1.208.grnaut\_07052003.113320.7333 -NCPV=6 -ICPV=3  
-NO\_XLPEX -NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N.Geneseq.101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	56	12.7	42	24	AAD27031	Human secreted epi
2	49	11.1	50	8	AA70336	3'-5' sequence of
3	47	10.6	51	22	AA178209	Human silent SNP c
4	47	10.6	60	24	ABN39530	Human spliced tran
5	46	10.4	60	24	ABN45376	Human spliced tran
6	46	10.4	65	24	ABN52611	Mouse spliced tran
7	46	10.4	66	24	ABN73405	Bovine embryonic g
8	46	10.4	68	20	AA335998	Polyrucleotide ass
9	46	10.4	71	21	AAA49245	HCV core protein p
10	46	10.4	71	21	AAA46147	Plasid pghCV/delt
11	46	10.4	72	14	AAQ51446	Human FACC Intron
12	46	10.4	79	14	AAQ36160	Mutagenic primer C
13	45	10.2	48	22	AA176614	Human silent SNP c
14	44.5	10.1	65	24	ABN31870	Rat spliced trans
15	44	10.0	50	15	AAQ54486	Excitatory amino a
16	44	10.0	50	15	AAQ54487	Excitatory amino a
17	44	10.0	60	21	AAA50222	Hybrid leptin/tiss
18	44	10.0	63	21	AAA50221	Hybrid leptin/tiss
19	44	10.0	75	22	AAQ94429	Virus-tpa vaccine
20	43.5	9.8	65	24	ABN31369	Rat spliced trans
21	43.5	9.8	76	21	AAA99033	Human TGC839 PCR p
22	43.5	9.8	80	24	ABL59337	PCR primer for hum
23	43	9.7	50	22	AA130844	Human SNP oligonuc
24	43	9.7	51	22	AA175717	Human silent SNP c
25	43	9.7	60	24	ABN38069	Human spliced tran
26	43	9.7	60	24	ABN40440	Human spliced tran
27	43	9.7	65	24	ABN31099	Rat spliced trans
28	43	9.7	66	21	AAA50949	DNA encoding tiss
29	43	9.7	66	21	AAA50218	Modified human tis
30	43	9.7	66	22	AA665513	Tissue plasminogen
31	43	9.7	69	21	AAA50217	Native human tissu
32	43	9.7	69	21	AAA50219	Modified human tis
33	43	9.7	69	21	AAA50220	Modified human tis
34	43	9.7	69	21	AAA509463	lPA leader sequenc
35	43	9.7	70	24	ABL39945	Signal peptide tpa
36	43	9.7	70	23	ABL58043	Cyclooxigenase-2,
37	43	9.7	75	24	ABL39944	Signal peptide tpa
38	43	9.7	75	21	AAA50338	Synthetic oligonuc
39	43	9.7	78	17	AA115476	Human tissue-speci
40	43	9.7	78	18	AA785512	Human tissue-speci
41	43	9.7	78	19	AAV21738	Sense oligomer 1 u
42	43	9.7	78	19	AAV21739	Antisense oligomer
43	43	9.7	78	22	AA509203	Synthetic sense ol
44	43	9.7	78	22	AA509204	Synthetic antisens
45	43	9.7	78	22	AAQ9430	Virus-LPA vaccine

## ALIGNMENTS

RESULT 1	
AAD27031	
ID	AAD27031 standard; DNA: 42 BP.
XX	
AC	AAD27031:
XX	
DT	09-APR-2002 (first entry)
XX	
DE	Human Secreted epithelial colon stromal-1 DNA amplifying PCR primer #5.
XX	
KW	secreted epithelial colon stromal-1; Sacs-1; gene therapy; osteoporosis;
KW	hematopoietic disorder; osteoporosis; osteogenesis imperfecta; cachexia;
KW	Paget's disease; periodontal disease; hypercalcaemia; glomerulonephritis;
KW	diabetes; obesity; osteopathic; cytostatic; nephrotrophic; antidiabetic;
KW	anorectic; immunomodulator; antipsoriatic; vulnery; antinfertility;
KW	gynaecological; antitumor; antinflammatory; cancer; cell therapy;
KW	human; PCR primer; ss.
XX	

```
OS Homo sapiens.
XX
XX MO200198497-A1.
XX
XX 27-DEC-2001.
XX
XX 28-NOV-2000; 2000MO-US32479.
XX
XX 21-JUN-2000; 2000US-0599087.
XX
XX 28-NOV-2000; 2000US-0724000.
XX
XX (AMGE-) AMGEN INC.
XX
XX Polverino AJ, Luethy R;
XX
XX WPI; 2002-122281/16.
XX
XX Secreted epithelial colon stromal-1 polypeptides and nucleic acids,
XX useful for diagnosing, treating and preventing hematopoietic disorder,
XX osteoporosis, Paget's disease, cancer, diabetes -
XX
XX Example 4; Page 89; 134pp; English.
XX
XX The present invention relates to an isolated murine or human secreted
XX epithelial colon stromal-1 (Secs-1) polypeptide, its allelic or splice
XX variant, orthologue, fragment or mutant. Secs-1 gene is used in gene
XX therapy and cell therapy. Secs-1 is useful for identifying a compound
XX which binds to a Secs-1 polypeptide. Secs-1 is useful for treating,
XX preventing or ameliorating a disease condition such as haematopoietic
XX disorder, osteoporosis, osteopetrosis, osteogenesis imperfecta, Paget's
XX disease, periodontal disease, hypercalcaemia, acute glomerulonephritis,
XX chronic glomerulonephritis, cancer, diabetes, obesity or cachexia. Secs-1
XX is also useful for diagnosing a pathological condition which involves
XX determining the presence or amount of Secs-1 or polypeptide encoded by
XX Secs-1 DNA in a sample; and diagnosing a pathological condition, or
XX susceptibility to pathological condition based on the presence or amount
XX of expression of the polypeptide. The present sequence is human Secs-1
XX DNA amplifying PCR primer.
XX
XX SQ Sequence 42 BP; 15 A; 10 C; 9 G; 8 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 224 Length: 42
XX Score: 56.00 Matches: 10
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 12.67% Indels: 0
XX DB: 24 Gaps: 0
XX
XX US-09-599-087B-5 (1-81) x AAN27031 (1-42)
XX
XX QY 25 LysArgArgProAlaIalysAlaItrPserGly 34
XX |||||||
XX DB 13 AAACGTGCTCCAGCTAAAGCCTGCTCAGGC 42
XX
XX RESULT 2
XX AAN70336/c
XX ID AAN70336 standard; DNA: 50 BP.
XX
XX AC AAN70336;
XX
XX 03-OCR-2002 (updated)
XX DT 02-APR-1991 (first entry)
XX
XX 3'-5' sequence of probe A11 used to detect hepatitis B virus (HBV)
XX analyse.
XX
XX Hepatitis B virus assay; diagnosis; ss.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX FH misc_feature 1..20
XX
```

```
FT /*tag- a
FT /note="Fluorescein label conjugate binding site"
FT misc_feature 21..50
FT /tag- b
FT /note="probe segment to HBV"
XX
XX EP225807-A.
XX
XX 16-JUN-1987.
XX
XX 10-DEC-1986; 86EP-0309622.
XX
XX 11-DEC-1985; 85US-0807624.
XX
XX 23-DEC-1986; 86US-0945876.
XX
XX (CHIR-) CHIRON CORP.
XX
XX Ureda MS, Warner B, Horn J;
XX
XX WPI; 1987-165004/24.
XX
XX Detecting specific nucleic acid sequences - by incubating sample
XX with labelling reagent set and capturing reagent set.
XX
XX Example; page 17; 31pp; English.
XX
XX In the example a BglII fragment of hepatitis B virus (AAN70325) was
XX detected. 12 sequences, complementary to different sequences
XX present in HBV were constructed (see AAN70326-n70337). Six of the HBV
XX complementary sequences are joined to a common sequence (A) for
XX complementing with the label conjugate (A'). The other six HBV
XX complementary sequences are joined to a common sequence (B) for
XX complementing with a biotinylated sequence (B') or a third DNA sequence
XX (B'') for binding to a support (see AAN70338, AAN70339, AAN70340).
XX (Updated on 03-OCR-2002 to add missing OS field.)
XX
XX SQ Sequence 50 BP; 13 A; 8 C; 14 G; 15 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 1.56e+03 Length: 50
XX Score: 49.00 Matches: 8
XX Percent Similarity: 78.57% Conservative: 3
XX Best Local Similarity: 57.14% Mismatches: 0
XX Query Match: 11.09% Indels: 0
XX DB: 8 Gaps: 0
XX
XX US-09-599-087B-5 (1-81) x AAN70336 (1-50)
XX
XX QY 40 CysCysHisArgValProSerProAsnSerThrAsnLeuLys 53
XX |||||||
XX DB 45 TGTTCATCATCAAGGCTTCACACACTTGACCACTTGCA 4
XX
XX RESULT 3
XX AAT78209
XX ID AAT78209 standard; DNA: 51 BP.
XX
XX AC AAT78209;
XX
XX 09-NOV-2001 (first entry)
XX
XX Human silent SNP containing nucleic acid SEQ:5150.
XX
XX Human: single nucleotide polymorphism; SNP; genome; gene therapy;
XX protein therapy; vaccine; probe; diagnostic assay; detection;
XX quantitation; restorative therapy; polymorphic; ds.
XX
XX Homo sapiens.
XX
XX MO200140521-A2.
XX
XX 07-JUN-2001.
XX
XX 30-NOV-2000; 2000MO-US32758.
XX
```

XX	30-NOV-1999;	99US-0168138.	
PR	29-NOV-2000;	2000US-0726173.	
XX	(CURA-) CURAGEN CORP.		
PA	ShlmKets RA, Leach M;		
PI	WPI; 2001-356160/37.		
XX			
DR			
PT	Polymorphic nucleic acid sequences, useful in genetic testing and		
PT	therapy -		
XX			
PS	Claim 1; Page 2086; 2653pp; English.		
XX			
CC	AAI73060 to AAI79867 represent isolated human polymorphic polynucleotide		
CC	sequences (I), which contain single nucleotide polymorphisms (SNPs).		
CC	AAI53114 to AAI53139 represent peptides related to human polymorphic		
CC	polynucleotide sequences. The sequences can be used in gene and protein		
CC	therapy, and in vaccine production. (I) and the polypeptides encoded by		
CC	them may be used in the prevention, diagnosis and treatment of diseases		
CC	associated with inappropriate expression of polymorphic polypeptides.		
CC	For example, (I) may be used to treat disorders by rectifying mutations		
CC	or deletions in a patient's genome that affect the activity of		
CC	polypeptides by expressing inactive proteins or to supplement the		
CC	patient's own production of polypeptide. Additionally, (I) and its		
CC	complementary sequences may also be used as DNA probes in diagnostic		
CC	assays to detect and quantitate the presence of similar nucleic acids		
CC	in samples, and therefore which patients may be in need of restorative		
CC	therapy. The polypeptides encoded by (I) may be used as antigens in the		
CC	production of antibodies specific for polymorphic polypeptides. The		
CC	antibodies may also be used to down regulate expression and activity.		
CC	The antibodies may also be used as diagnostic agents for detecting the		
CC	presence of polymorphic polypeptides in samples.		
XX			
SO	Sequence 51 BP; 14 A; 13 C; 13 G; 11 T; 0 other;		
Alignment Scores:			
Pred. NO.:	2.63e+03	Length:	51
Score:	47.00	Matches:	8
Percent Similarity:	72.73%	Conservative:	0
Best Local Similarity:	72.73%	Mismatches:	3
Query Match:	10.63%	Indels:	0
DB:	22	Gaps:	0
US-09-599-087B-5 (1-81) x AAI78209 (1-51)			
OY	39 LeuCYSCYSHsArgValProSerProAsnSer 49		
Db	19 TTGTGCTGTGAACGACGACCGCTCAGAGAAATTC 51		
RESULT 4			
ID	ABN39530/c		
XX	ABN39530 standard; DNA: 60 BP.		
AC	ABN39530;		
XX			
DT	15-JUL-2002 (first entry)		
XX			
DE	Human spliced transcript detection oligonucleotide SEQ ID NO:12278.		
XX			
KW	Human; mouse; rat; splice transcript; detection; RNA transcript;		
OS	splice variant; transcriptome; oligonucleotide library; ss.		
XX			
OS	Homo sapiens.		
XX			
PD	WO200210449-A2.		
XX			
PD	07-FEB-2002.		
XX			
PF	20-JUL-2001; 2001MO-IB01903.		
XX			
PR	28-JUL-2000; 2000US-221607P.		

```

PR 02-MAY-2001: 2001US-287724P.
XX
XX (COMP-) COMPUGEN INC.
XX
XX Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX
XX WPI; 2002-257383/30.
XX
XX
XX New oligonucleotide libraries comprising oligonucleotides which
XX selectively hybridize to mRNAs transcribed from a transcription unit of
XX a genome, useful for detecting tissue-, pathology-, and
XX developmental-specific genes
XX
XX Example 1; SEQ ID 12278; 47pp; English:
XX
XX The present invention describes oligonucleotide libraries for detecting
XX messenger RNAs that populate a (sub-)transcriptome, where the
XX (sub-)transcriptome comprises messenger RNAs transcribed from multiple
XX transcription units that populate a genome. The library comprises
XX several oligonucleotides, each capable of hybridizing selectively to a
XX set of messenger RNAs transcribed from a given transcription unit of
XX the genome, which encodes one or more messenger RNA splice variants.
XX The oligonucleotide libraries are useful for detecting mRNAs from a
XX biological sample, in expression profiling studies, in qualitatively or
XX quantitatively characterizing the corresponding transcriptome, and in
XX detecting RNA transcripts and splice variants of human or animal
XX transcriptomes. The libraries may also be used as specialised mini
XX libraries to detect transcripts of a sub-transcriptome under a
XX particular biological or pathological state, and so allowing the
XX detection of tissue- and pathology-specific genes such as those genes
XX only expressed in specific tissue under a specific pathological
XX condition; to detect developmental specific genes; and to detect RNA
XX transcripts and splice variants of a transcriptome of a patient suffering
XX from a particular disorder. ABN27253 to ABN59589 represent
XX oligonucleotide sequences from rats, humans and mice, which are used in
XX the exemplification of the present invention.
XX N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 60 BP; 13 A; 23 C; 9 G; 15 T; 0 other;
XX
XX Alignment Scores:
XX Pred. NO.: 3.23e+03 Length: 60
XX Score: 47.00 Matches: 9
XX Percent Similarity: 71.43% Conservative: 1
XX Best local Similarity: 64.29% Mismatches: 4
XX Query Match: 10.63% Indels: 0
XX DB: 24 Gaps: 0
XX
XX US-09-599-087B-5 (1-81) x ABN39530 (1-60)
XX
XX Oy 21 serthrgluclylsyargarproalatalysalatrpsergly 34
XX |||:::||||| ||||||| |||||||
XX Db 56 TCGTCTGAAGCCTAAGGAGACCTTGCGTTGATGAGATGGG 15
XX
XX RESULT 5
XX ID ABN45376/c ABN45376/c
XX AC ABN45376;
XX
XX 15-JUL-2002 (first entry)
XX
XX Human spliced transcript detection oligonucleotide SEQ ID NO:18124.
XX
XX Human; mouse; rat; splice transcript; detection; RNA transcript;
XX splice variant; transcriptome; oligonucleotide library; ss.
XX
XX Homo sapiens.
XX
XX WO200210449-A2.
XX
XX

```

```

PD 07-FEB-2002.
PP
PE 20-JUL-2001; 2001MO-B01903.
PR 28-JUL-2000; 2000US-221607P.
PR 02-MAY-2001; 2001US-287724P.
XX
XX (COMP-) COMUGEN INC.
PA Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
PI WPI; 2002-257383/30.
XX
XX New oligonucleotide libraries comprising oligonucleotides which
PT selectively hybridize to mRNAs transcribed from a transcription unit of
PT a genome, useful for detecting tissue-, pathology-, and
PT developmental-specific genes
XX
XX Example 1; SEQ ID 18124; 47pp; English.
XX
XX The present invention describes oligonucleotide libraries for detecting
CC messenger RNAs that populate a (sub-)transcriptome, where the
CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
CC transcription units that populate a genome. The library comprises
CC several oligonucleotides, each capable of hybridizing selectively to a
CC set of messenger RNAs transcribed from a given transcription unit of
CC the genome, which encodes one or more messenger RNA splice variants.
CC The oligonucleotide libraries are useful for detecting mRNAs from a
CC biological sample, in expression profiling studies, in qualitatively or
CC quantitatively characterizing the corresponding transcriptome, and in
CC detecting RNA transcripts and splice variants of human or animal
CC transcriptomes. The libraries may also be used as specialised mini
CC libraries to detect transcripts of a sub-transcriptome under a
CC particular biological or pathological state, and so allowing the
CC detection of tissue- and pathology-specific genes such as those genes
CC only expressed in specific tissue under a specific pathological
CC condition: to detect developmental specific genes; and to detect RNA
CC transcripts and splice variants of a transcriptome of a patient suffering
CC from a particular disorder. ABN27253 to ABN59589 represent
CC oligonucleotide sequences from rats, humans and mice, which are used in
CC the exemplification of the present invention.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 60 BP; 14 A; 16 C; 21 G; 9 T; 0 other;
SQ
XX
XX Alignment Scores:
XX Pred. No.: 4.13e+03 Length: 60
XX Score: 46.00 Matches: 7
XX Percent Similarity: 63.64 Conservative: 0
XX Best Local Similarity: 63.64 Mismatches: 4
XX Query Match: 10.41 Indels: 0
XX DB: 24 Gaps: 0
XX
XX US-09-599-087B-5 (1-81) x ABN45376 (1-60)
XX
XX QY 62 ProCysLysLeuGluProGluProArgLeuTrp 72
XX ||||| ||| ||| ||||| |||
XX DB 55 CCGTGACGCTCGTCGTCGACGACCGAATCTCTG 23
XX
XX RESULT 6
XX ID ABN52611/c
XX ID ABN52611 standard; DNA; 65 BP.
XX
XX ABN52611:
XX
XX 15-JUL-2002 (first entry)
XX
XX Mouse spliced transcript detection oligonucleotide SEQ ID NO:25359.
XX
XX Human: mouse: rat: splice transcript: detection; RNA transcript;
XX splice variant; transcriptome; oligonucleotide library; ss.
XX

```

```

XX  Mus musculus.
OS
XX
XX  WO200210449-A2.
PN
XX
XX  07-FEB-2002.
PD
XX
XX  20-JUL-2001; 2001WO-1B01903.
PF
XX
XX  28-JUL-2000; 2000US-221607P.
XX
PR  02-MAY-2001; 2001US-287724P.
XX
XX  (COMP-) COMPUSEN INC.
PA
XX
XX  Shoshan A, Wasserman A, Mintz E, Mintz L, Falgler S;
PI
XX
XX  WPI; 2002-257383/30.
DR
XX
XX  New oligonucleotide libraries comprising oligonucleotides which
PT  selectively hybridize to mRNAs transcribed from a transcription unit of
PT  a genome, useful for detecting tissue-, pathology-, and
PT  developmental-specific genes
XX
XX  Example 1; SEQ ID 25359; 47pp; English.
PS
XX
XX  The present invention describes oligonucleotide libraries for detecting
CC  messenger RNAs that populate a (sub-)transcriptome, where the
CC  (sub-)transcriptome comprises messenger RNAs transcribed from multiple
CC  transcription units that populate a genome. The library comprises
CC  several oligonucleotides, each capable of hybridizing selectively to a
CC  set of messenger RNAs transcribed from a given transcription unit of
CC  the genome, which encodes one or more messenger RNA splice variants.
CC  The oligonucleotide libraries are useful for detecting mRNAs from a
CC  biological sample, in expression profiling studies, in qualitatively or
CC  quantitatively characterizing the corresponding transcriptome, and in
CC  detecting RNA transcripts and splice variants of human or animal
CC  transcriptomes. The libraries may also be used as specialised mini
CC  libraries to detect transcripts of a sub-transcriptome under a
CC  particular biological or pathological state, and so allowing the
CC  detection of tissue- and pathology-specific genes such as those genes
CC  only expressed in specific tissue under a specific pathological
CC  condition: to detect developmental specific genes; and to detect RNA
CC  transcripts and splice variants of a transcriptome of a patient suffering
CC  from a particular disorder. ABN277253 to ABN35589 represent
CC  oligonucleotide sequences from rats, humans and mice, which are used in
CC  the exemplification of the present invention.
CC  N.B. The sequence data for this patent did not form part of the printed
CC  specification, but was obtained in electronic format directly from WIPO
CC  at ftp.wipo.int/pub/published_pcr_sequences.
XX
XX  Sequence 65 BP; 20 A; 16 C; 13 G; 16 T; 0 other;
SO
XX
XX  Alignment Scores:
XX
XX  Pred. NO.: 4.58e+03 Length: 65
XX
XX  Score: 46.00 Matches: 8
XX
XX  Percent Similarity: 78.57% Conservative: 3
XX
XX  Best Local Similarity: 57.14% Mismatches: 3
XX
XX  Query Match: 10.41% Indels: 0
XX
XX  DB: 24 Gaps: 0
XX
XX  US-09-599-087B-5 (1-81) x ABN52611 (1-65)
XX
XX  QY 38 ArgLeuCySCyHisArgValProSerProAsnSerThrAsn 51
XX  |||:||||| |||||:||||| |||
XX  DB 61 AGAGTTTGCACTCATAGATCCCAAGTGGTCTTCAATAAAT 20
XX
XX  RESULT 7
XX  ID ABN73405 standard; cDNA; 66 BP.
XX
XX  ABN73405;
XX
XX  03-JUL-2002 (first entry)
XX

```



```

XX PA (MED-) MEDICAL RES COUNCIL.
XX PI Hope G, McLauchlan J;
XX DR WPI; 2000-400040/34.
XX PT Protein comprising a lipid globule targeting sequence consisting of a
XX PT hepatitis C virus core protein, useful for targeting a protein of
XX PT interest to lipid globules which are subsequently secreted into animal
XX PS milk.
XX PS Example; Page 32; 60pp; English.
XX CC The present sequence is a mutagenic primer used in the construction of a
XX CC plasmid containing a mutated version of the hepatitis C virus core
XX CC protein gene. The core protein has been shown to be localised to lipid
XX CC droplets within cells. This means that fragments of the protein, in
XX CC particular those comprising amino acids 125-144 or 161-166, can be used
XX CC to target proteins to lipid droplets in milk. This provides a simple and
XX CC efficient means of obtaining proteins of interest, using transgenic
XX CC animals. Proteins of interest include those implicated in disease and
XX CC antigenic polypeptides for use as vaccines.
XX SQ Sequence 71 BP; 13 A; 20 C; 22 G; 16 T; 0 other;

Alignment Scores:
Pred. No.: 5.13e+03 Length: 71
Score: 46.00 Matches: 9
Percent Similarity: 68.75% Conservative: 2
Best Local Similarity: 56.25% Mismatches: 5
Query Match: 10.41% Indels: 0
DB: 21 Gaps: 0

US-09-599-087b-5 (1-81) x AAA49245 (1-71)
OY 25 LYSARGARGPROALALysAlATrPSerGLyARgARgThrARgucys 40
DB 19 CGTGGGCGCCCTCTGTGTCGGGTTCTGGAAGACGGTGACTATGC 66

RESULT 10
AAA46147
ID AAA46147 standard; cDNA; 71 BP.
XX
AC AAA46147;
XX
DT 27-OCT-2000 (first entry)
XX
DE Plasmid pGHCV/delta145-154 Insertion oligonucleotide.
XX
KM Hepatitis C virus; core protein; HCV; lipid globule; chronic hepatitis;
XX KM liver disease; ADRP displacement; Insertion oligonucleotide; ss.
XX OS Synthetic.
XX PN WO200031532-A1.
XX PD 02-JUN-2000.
XX PF 23-NOV-1999; 99WO-GH03906.
XX PR 26-NOV-1998; 98GB-0025951.
XX PA (MED-) MEDICAL RES COUNCIL.
XX PI Hope G, McLauchlan J;
XX DR WPI; 2000-400181/34.
XX PT Identifying a substance for treating or preventing a viral infection
XX PT such as hepatitis C virus, comprises determining if the substance
XX PT disrupts a lipid globule target sequence to lipid globule interaction

```

```

XX PS Examples; Page 37; 74pp; English.
XX CC The present sequence is an oligonucleotide which was inserted into
XX CC a plasmid which contained part of hepatitis C virus core protein gene.
XX CC The virus causes chronic hepatitis and liver disease in humans. It is
XX CC thought that it does this by associating with intracellular lipid
XX CC droplets and downregulating the expression of adipocyte-specific
XX CC differentiation-related protein (ADRP), which is probably required for
XX CC the maintenance of lipid droplets. The core protein can be used to
XX CC identify substances capable of disrupting this interaction, which are
XX CC thus capable of treating or preventing infection by the hepatitis C virus
XX CC and its accompanying effects. This oligonucleotide was used in the
XX CC construction of mutants of the core protein, which were used to determine
XX CC their efficiency of infection compared to the wild-type.
XX SQ Sequence 71 BP; 13 A; 20 C; 22 G; 16 T; 0 other;

Alignment Scores:
Pred. No.: 5.13e+03 Length: 71
Score: 46.00 Matches: 9
Percent Similarity: 68.75% Conservative: 2
Best Local Similarity: 56.25% Mismatches: 5
Query Match: 10.41% Indels: 0
DB: 21 Gaps: 0

US-09-599-087b-5 (1-81) x AAA46147 (1-71)
OY 25 LYSARGARGPROALALysAlATrPSerGLyARgARgThrARgucys 40
DB 19 CGTGGGCGCCCTCTGTGTCGGGTTCTGGAAGACGGTGACTATGC 66

RESULT 11
AAQ51446/C
ID AAQ51446 standard; DNA; 72 BP.
XX
AC AAQ51446;
XX
DT 20-MAY-1994 (first entry)
XX
DE Human FACC Intron 9 3' region.
XX
KM Fanconi Anemia Group C; FACC; complementing cDNA; variant; diagnosis;
XX KM open reading frame; Fanconi anemia; gene therapy; ss.
XX OS Homo sapiens.
XX PN WO9322435-A.
XX PD 11-NOV-1993.
XX PF 27-APR-1993; 93WO-CA00178.
XX PR 29-APR-1992; 92US-0876285.
XX PR 21-JUL-1992; 92US-0918313.
XX PR 15-JAN-1993; 93US-0003963.
XX PA (HOSP-) HOSPITAL FOR SICK CHILDREN.
XX PA (UNME-) UNITED MEDICAL & DENTAL SCHOOL GUYS.
XX PI Buchwald M, Mathew CG, Strathdee CA, Wevrick R;
XX DR WPI; 1993-368794/46.
XX PT Human cDNA which complements Fanconi Anaemia gp. C - used to
XX PT develop prods. for use in diagnosis, study and therapy of Fanconi
XX PT Anaemia
XX PS Claim 1; Page 111; 137pp; English.
XX CC The sequences given in AAQ51429-54 represent the 5' and 3' splice
XX CC regions of the introns from the Fanconi Anemia Group C Complementing
XX CC (FACC) DNA. Three cDNA molecules which are cellular variants of a

```

CC single cDNA, are transcribed from the FACC gene. The three cDNAs  
 CC each contain an identical open reading frame encoding the FACC  
 CC protein. FACC protein may be used for the diagnosis and study of  
 CC Fanconi anemia. The FACC gene and cDNAs may be used in gene therapy.  
 XX

SQ Sequence 72 BP; 12 A; 16 C; 13 G; 31 T; 0 other;

#### Alignment Scores:

Pred. No.:	Length:	Matches:	Score:
5.22e+03	72	7	46.00
Percent Similarity:	76.92%	Conservative:	3
Best Local Similarity:	53.85%	Mismatches:	3
Query Match:	10.41%	Indels:	0
DB:	14	Gaps:	0

US-09-599-087b-5 (1-81) x AAQ51446 (1-72)

OY 41 CysHisArgValProSerProAsnSerThrAsnLeuLys 53  
 |||||  
 DB 71 TGCCACGAGATGAAATCCAAAGACATGACATTAAAG 33

#### RESULT 12

AAQ36160  
 ID AAQ36160 standard; DNA; 79 BP.

AC AAQ36160;

DT 27-MAY-1993 (first entry)

DE Mutagenic primer COD1172.

KW Mutagenesis; site specific; nucleic acid constructs;  
 restriction site; introduction; removal; ss.

OS Synthetic.

PN W09301282-A.

PD 21-JAN-1993.

PF 01-JUL-1992; 92WO-US05573.

PR 01-JUL-1991; 91US-0724237.

PA (BERL-) BERLEX LAB INC.

PI Andrews WH, Morser MJ, Vliander LR.

DR WPI; 1993-045488/05.

PT Site-specific mutagenesis of nucleic acid constructs - using an  
 oligonucleotide which changes a nucleotide and introduces or  
 removes a restriction site

PS Example; Page 48; 87pp; English.

CC The sequence is that of mutagenic primer COD1172 which was used as part  
 CC of a method for site-specific mutagenesis of nucleic acid constructs.  
 CC The primer changes a nucleotide and introduces or removes a restriction  
 CC site. Using the method multiple mutations can be performed on a single  
 CC target sequence simultaneously and more than one such target sequence  
 CC can be included on a single construct. The method allows screening of  
 CC mutants simply by restriction analysis. The primer was used in the  
 CC construction of mutagenesis plasmid pMW3.0-19, it was annealed with  
 CC COD1169, COD1170, COD1171 and COD1173 and inserted into the  
 CC EcoRI-HindIII sites of pUC18.  
 CC

SQ Sequence 79 BP; 10 A; 27 C; 32 G; 10 T; 0 other;

Alignment Scores:  
 Pred. No.: 5.88e+03 Length: 79  
 Score: 46.00 Matches: 8  
 Percent Similarity: 52.94% Conservative: 1

Best Local Similarity: 47.06% Mismatches: 8  
 Query Match: 10.41% Indels: 0  
 DB: 14 Gaps: 0

US-09-599-087b-5 (1-81) x AAQ36160 (1-79)

OY 29 AlAlysAlaTrpSerGjArGArGhrArgLeuGcysGysHisArgValPro 45  
 |||  
 DB 15 GCGTCGGGCTGGCGGGGGCTGCCAGAGCCGCTGTGCCACCGCCACT 65

#### RESULT 13

AA176614/c  
 ID AA176614 standard; DNA; 48 BP.

AC AA176614;

DT 09-NOV-2001 (first entry)

DE Human silent SNP containing nucleic acid SEQ:3555.

KW Human; single nucleotide polymorphism; SNP; genome; gene therapy;  
 protein therapy; vaccine; probe; diagnostic assay; detection;  
 quantitation; restorative therapy; polymorphic; ds.

OS Homo sapiens.

PN W0200140521-A2.

PD 07-JUN-2001.

PF 30-NOV-2000; 2000WO-US32758.

PR 30-NOV-1999; 99US-0168138.

PR 29-NOV-2000; 2000US-0726173.

PA (CURA-) CURAGEN CORP.

PI Shimkets RA, Leach M;

DR WPI; 2001-356160/37.

PT Polymorphic nucleic acid sequences, useful in genetic testing and  
 therapy -

PS Claim 1; Page 1139; 2653pp; English.

CC AA173060 to AA179867 represent isolated human polymorphic polynucleotide  
 CC sequences (I), which contain single nucleotide polymorphisms (SNPs).  
 CC AA173114 to AA173129 represent peptides related to human polymorphic  
 CC polynucleotide sequences. The sequences can be used in gene and protein  
 CC therapy, and in vaccine production. (I) and the polypeptides encoded by  
 CC them may be used in the prevention, diagnosis and treatment of diseases  
 CC associated with inappropriate expression of polymorphic polypeptides.  
 CC For example, (I) may be used to treat disorders by rectifying mutations  
 CC or deletions in a patient's genome that affect the activity of  
 CC polypeptides by expressing inactive proteins or to supplement the  
 CC patients own production of polypeptide. Additionally, (I) and its  
 CC complementary sequences may also be used as DNA probes in diagnostic  
 CC assays to detect and quantitate the presence of similar nucleic acids  
 CC in samples, and therefore which patients may be in need of restorative  
 CC therapy. The polypeptides encoded by (I) may be used as antigens in the  
 CC production of antibodies specific for polymorphic polypeptides. The  
 CC antibodies may also be used to down regulate expression and activity.  
 CC The antibodies may also be used as diagnostic agents for detecting the  
 CC presence of polymorphic polypeptides in samples.  
 CC

SQ Sequence 48 BP; 8 A; 15 C; 20 G; 5 T; 0 other;

Alignment Scores:  
 Pred. No.: 3.97e+03 Length: 48  
 Score: 45.00 Matches: 8  
 Percent Similarity: 76.92% Conservative: 2  
 Best Local Similarity: 61.54% Mismatches: 3

```

Query Match.: 10.18% Indels: 0
DB: 22 Gaps: 0
US-09-599-087B-5 (1-81) x AAI76614 (1-48)
Qy 67 ProGluProArGLeuTrpValValProGlyAlaLeuPro 79
   ||| |||||
   ::|||
Db 44 CCAGATCCCGACTCCTTGGCCTCCAGAGGCCATCCG 6

RESULT 14
ID ABN31870 standard; DNA; 65 BP.
AC ABN31870:
XX
XX
XX 15-JUL-2002 (first entry)
XX
XX
XX Rat spliced transcript detection oligonucleotide SEQ ID NO:4618.
XX
XX
XX Human: mouse; rat; splice transcript; detection; RNA transcript;
XX splice variant; transcriptome; oligonucleotide library; ss.
XX
XX Rattus norvegicus.
XX
XX MO200210449-A2.
XX
XX PD 07-FEB-2002.
XX
XX 20-JUL-2001; 2001WO-IB01903.
XX
XX PF 28-JUL-2000; 2000US-221607P.
XX PR 02-MAY-2001; 2001US-287724P.
XX
XX PA (COMP-) COMPUGEN INC.
XX
XX PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX
XX WPI; 2002-257383/30.
XX
XX PT New oligonucleotide libraries comprising oligonucleotides which
XX selectively hybridize to mRNAs transcribed from a transcription unit of
XX a genome, useful for detecting tissue-, pathology-, and
XX PT developmental-specific genes
XX
XX PS Example 1; SEQ ID 4618; 47pp; English.
XX
XX
XX The present invention describes oligonucleotide libraries for detecting
XX messenger RNAs that populate a (sub-)transcriptome, where the
XX (sub-)transcriptome comprises messenger RNAs transcribed from multiple
XX CC transcription units that populate a genome. The library comprises
XX several oligonucleotides, each capable of hybridising selectively to a
XX set of messenger RNAs transcribed from a given transcription unit of
XX the genome, which encodes one or more messenger RNA splice variants.
XX The oligonucleotide libraries are useful for detecting mRNAs from a
XX biological sample. In expression profiling studies, in qualitatively or
XX quantitatively characterising the corresponding transcriptome, and in
XX CC detecting RNA transcripts and splice variants of human or animal
XX transcripts. The libraries may also be used as specialised mini
XX libraries to detect transcripts of a sub-transcriptome under a
XX particular biological or pathological state, and so allowing the
XX detection of tissue- and pathology-specific genes such as those genes
XX only expressed in specific tissue under a specific pathological
XX condition; to detect developmental specific genes; and to detect RNA
XX transcripts and splice variants of a transcriptome of a patient suffering
XX from a particular disorder. ABN27253 to ABN93589 represent
XX oligonucleotide sequences from rats, humans and mice, which are used in
XX CC the exemplification of the present invention.
XX N.B. The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pt_sequences.
XX
XX Sequence 65 BP; 19 A; 20 C; 13 G; 13 T; 0 other;

```

Alignment Scores:		Length:	
Pred. No.:	6.63e+03	Matches:	65
Score:	44.50	Conservative:	9
Percent Similarity:	70.59%	Mismatches:	3
Best Local Similarity:	52.94%	Indels:	2
Query Match:	10.07%	Gaps:	1
DB:	24		

  

US-09-599-087B-5 (1-81) x ABBN1870 (1-65)	
QY	43 ArgValProSerProAsnSer-----ThaAsnLeuTysGlyHisHis 56
DB	1 AGAATTCAGAACCCCTCGTCCAAAGAGTGCACCAACACCTTCAGAGTCACTAT 51

  

RESULT 15	
ID	AA054486/c
XX	AA054486 standard; cDNA; 50 BP.
XX	AA054486;
XX	07-JUL-1994 (first entry)
XX	
XX	Excitatory amino acid receptor (EAA4a) coding sequence fragment.
DE	
XX	Excitatory amino acid receptor; EAA4a; screening; detection;
KW	identification; CNS; nervous disorders; antibody; oocyte; membrane
KM	protein; ion-channel; therapeutics; ss.
XX	
OS	Homo sapiens.
XX	
PN	EP578409-A.
XX	
XX	12-JAN-1994.
XX	
XX	24-JUN-1993; 93EP-0304961.
PR	
XX	24-JUN-1992; 92US-0903456.
XX	
PA	(ELLI/) ELLIOTT C E.
PA	(KAMB/) KAMBOJ R.
PA	(NOTT/) NOTT S L.
PI	Kamboj R, Nutt SL, Elliott CE;
PI	WPI; 1994-010248/02.
DR	
XX	
PT	Excitatory amino acid for receptors with kainate binding activity
PT	- used to test ligands for CNS receptor interaction to identify
PT	cpds. useful against CNS disorders
XX	
PS	Disclosure; Figure 3b; 37pp: English.
XX	
XX	The excitatory amino acid receptor (See AAN45367) can be used to
CC	screen ligands. The ligands can be assayed for interaction by
CC	incubating them with cells, preferably oocytes, expressing the
CC	receptor or with membrane preparations from these cells, and
CC	assessing any interaction by determining receptor-ligand binding or
CC	ligand-mediated ion channel activation. This method is useful in
CC	the identification of therapeutics useful to treat CNS disorders in
CC	humans. Fragments of the receptor are useful for structural
CC	investigations and to raise antibodies. This fragment of the coding
CC	sequence (AA054482) corresponds to bases 1971 to 2020.
XX	
SO	Sequence 50 BP; 11 A; 11 C; 15 G; 13 T; 0 other;

  

Alignment Scores:		Length:	
Pred. No.:	5.35e+03	Matches:	50
Score:	44.00	Conservative:	7
Percent Similarity:	60.00%	Mismatches:	2
Best Local Similarity:	46.67%	Indels:	0
Query Match:	9.95%	Gaps:	0
DB:	15		



Fri May 9 16:38:54 2003

us-09-599-087b-5\_1.rng

Page 9

US-09-599-087B-5 (1-81) x AAQ54486 (1-50)

QY 34 G1yArGArGThrArGLeucYScYSH1sArgYAlProSerProasn 48  
|||  
::: ||||| |||  
Db 46 GGcATGAGcTCAGAcCTGcTGcATGAGAGcTCcCAAcTCcCAAc 2

Search completed: May 8, 2003, 21:52:57  
Job time : 210 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 8, 2003, 21:48:43 ; Search time 62 seconds

400.658 million cell updates/sec

Title: US-09-599-087B-5

Sequence: 1 MRLVLSSLLCILLCFSTF.....PCKLEPEPRRLWVPGALPQV 81

	BLOSUM62	yearnext
Scoring table:	0	5
yearcod	10	0

Xgapop	10.0	Xgapext	0.5
Ygapop	10.0	Ygapext	0.5
Fgapop	6.0	Fgapext	7.0
Delop	6.0	Delext	7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 322676

```

Minimum DB seq length: 25
Maximum DB seq length: 80

```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Command line parameters:

```

0+/-cgn2.1/USP0.spool/US05959087/runtat07052003.113322.7375/arp_query.fasta.1.1263
DB-assigned.patients.NA.-OBMT-fastap.-SOFPIX-rmi1.MINMATCH-0.1.-LOOPEC=0
-LOOPEC=0 -UNITS-bits -START-1 -END-1 -MATRIX-blosum62 -TRANS-numa40.ccd1
-LIST-45 -POCALIGN-200 -THR SCORE-Pct -THR MAX-0 -THR MIN-0 -ALIGN-15
-MODE-LOCAL -OUTFMT-Pto -NORM-ext HEMPSTAT-500 -MINLEN-25 -MAXLEN-80
-USER-US05959087_EGCG.1.1.36.6tunat.07052003.113322.7375 -NCPU-6 -ICPU-3
-NO_ALXPY -NO_MMAP -LARGESIZE -NEG_SCORES-0 -WAIT -DSBLOCK-100 -LONGLONG
-DEVTIMEOUT-120 -WARN_TIMEOUT-30 -THREADS-1 -XCAPO-10 -XGAPEXT-0.5 -FCAPO-6
-XGAPEXT-7 -YGAPO-10 -YGAPEXT-0.5 -DELPO-6 -DELEXT-7

```

Database : Issued\_Patents\_NA:\*

```

1: /cgn2_6/ptodata/2/1na/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/1na/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/1na/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/1na/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/1na/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/2/1na/backfiles1.seq: *

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	53	12.0	79	1	US-08-170-290A-19	Sequence 19, Appl
2	46.5	10.5	48	2	US-08-350-260A-289	Sequence 289, Appl
3	46	10.4	72	1	US-08-441-430-32	Sequence 32, Appl
c	45.5	10.3	69	1	US-08-554-612C-37	Sequence 37, Appl
5	45	10.2	39	2	US-08-350-260A-233	Sequence 233, Appl
6	44	10.0	42	1	US-08-662-255-23	Sequence 23, Appl
7	44	10.0	50	1	US-07-903-456-7	Sequence 7, Appl1
c	44	10.0	50	1	US-07-903-456-8	Sequence 8, Appl1
c	44	10.0	50	4	US-08-249-241-7	Sequence 7, Appl1
c	9	44	50	4	US-08-249-241-8	Sequence 8, Appl1
c	10	44	50	4	US-08-350-260A-287	Sequence 287, Appl
11	44	10.0	51	2	US-08-350-260A-287	Sequence 287, Appl
12	44	10.0	51	2	US-08-350-260A-290	Sequence 290, Appl

13	4.3	9.7	72	6	5466668-7
14	3.38	9.7	78	1	US-08-338-992B-5
15	4.3	9.7	78	1	US-08-338-992B-6
16	4.3	9.7	78	4	US-09-010-733-5
17	4.3	9.7	78	4	US-09-010-733-6
18	4.3	9.7	78	5	PCT-US95-09057-5
19	4.3	9.7	78	5	PCT-US95-09057-6
20	4.3	9.7	79	1	US-08-571-983-2
21	4.3	9.7	79	5	PCT-US94-08111-2
22	4.2	9.5	30	1	US-08-338-471D-4
23	4.2	9.5	69	1	US-07-937-284A-20
24	4.2	9.5	69	2	US-08-236-428B-20
25	4.1.5	9.4	46	4	US-08-872-055-19
26	4.1.5	9.4	60	2	US-08-465-380-112
27	4.1.5	9.4	60	2	US-08-480-478-27
28	4.1.5	9.4	60	2	US-08-486-397-112
29	4.1.5	9.4	60	2	US-08-486-399-112
30	4.1.5	9.4	60	2	US-08-461-965-112
31	4.1.5	9.4	60	2	US-08-326-110A-27
32	4.1.5	9.4	60	2	US-08-634-64-112A
33	4.1.5	9.4	60	3	US-09-249-4-71-112
34	4.1.5	9.4	60	3	US-09-249-4-71-112
35	4.1.5	9.4	60	3	US-09-249-4-45-112
36	4.1.5	9.4	60	3	US-08-809-4-45-112
37	4.1.5	9.4	60	3	US-09-249-4-45-112
38	4.1.5	9.4	60	3	US-09-249-4-46-112
39	4.1	9.3	30	1	US-07-969-931-27
40	4.1	9.3	30	1	US-07-855-417A-27
41	4.1	9.3	60	4	US-09-339-910A-13
42	4.1	9.3	60	4	US-09-339-910A-13
43	4.1	9.3	60	4	US-08-769-062B-13
44	4.1	9.3	60	4	US-09-344-062B-13
45	4.1	9.3	60	4	US-09-559-566C-13

## ALIGNMENTS

RESULT 1  
 US-08-170-290A-19  
 Sequence 19, Application US/08170290A  
 Patent No. 5702931  
 GENERAL INFORMATION:  
 APPLICANT: Andrews, William H.  
 APPLICANT: Morser, Michael J.  
 APPLICANT: Ziehlender, Laura R.  
 TITLE OF INVENTION: No. 5702931el Mutagenesis Methods and  
 TITLE OF INVENTION: Compositions  
 NUMBER OF SEQUENCES: 63  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: James M. Heslin  
 STREET: 379 Lytton Ave.  
 CITY: Palo Alto  
 STATE: California  
 COUNTRY: USA  
 ZIP: 94301  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/170,290A  
 FILING DATE: 28-DEC-1993  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US92/05573  
 FILING DATE: 01-JUL-1992  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/724,237  
 FILING DATE: 01-JUL-1991  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 ;  
 ;  
 ;

NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 11972-58-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 79 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-170-290A-19

Alignment Scores:  
Pred. No.: 114 Length: 79  
Score: 53.00 Matches: 9  
Percent Similarity: 55.00% Conservative: 2  
Best Local Similarity: 45.00% Mismatches: 0  
Query Match: 11.99% Indels: 0  
DB: 1 Gaps: 0

US-09-599-087B-5 (1-81) x US-08-170-290A-19 (1-79)

QY 29 AlAlysAlArpserGlyArGArGThrArgLeuCyScySHisArgValProserProAsn 48  
Db 15 GCGTCGGCGCTGGCGGGGCGCGCCAGAGCGCGTGTGCGCCACCGCTCCCGAGT 74

RESULT 2  
US-08-350-260A-289  
Sequence 289, Application US/08350260A  
Patent No. 5962255  
GENERAL INFORMATION:  
APPLICANT: Winter, Gregory Paul  
APPLICANT: Griffiths, Andrew David  
APPLICANT: Williams, Samuel Cameron  
APPLICANT: Waterhouse, Peter  
APPLICANT: Nissim, Aluva  
APPLICANT: Johnson, Kevin Stuart  
TITLE OF INVENTION: Methods for producing members of specific  
NUMBER OF SEQUENCES: binding pairs  
TITLE OF INVENTION: binding pairs  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David W. Clough  
STREET: Marshall, O'Toole, Gerstein, Murray & Borun  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/350, 260A  
FILING DATE: 05-DEC-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9110549.4  
FILING DATE: 15-MAY-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9206318.9  
FILING DATE: 24-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB91/01134  
FILING DATE: 10-JUL-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB92/00883  
FILING DATE: 15-MAY-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB93/00605  
FILING DATE: 24-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/150, 002  
FILING DATE: 31-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/307, 619  
FILING DATE: 16-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Clough, David W  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 28111/32372  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
INFORMATION FOR SEQ ID NO: 289:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 48 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-350-260A-289

Alignment Scores:  
Pred. No.: 320 Length: 48  
Score: 46.50 Matches: 9  
Percent Similarity: 66.67% Conservative: 3  
Best Local Similarity: 50.00% Mismatches: 3  
Query Match: 10.52% Indels: 3  
DB: 2 Gaps: 1

US-09-599-087B-5 (1-81) x US-08-350-260A-289 (1-48)

QY 32 TrpserGlyArGArGThrArgLeuCyScySHisArgValProserProAsnSer 49  
Db 3 TGCTCCCTCCGCCGAGATACAGTGTCTGT-----CCACGCGCTGACGT 47

RESULT 3  
US-08-441-430-22/C  
Sequence 22, Application US/08441430  
Patent No. 5681942  
GENERAL INFORMATION:  
APPLICANT: Buchwald, Manuel  
APPLICANT: Strathdee, Craig A.  
APPLICANT: Weyrick, Rachel  
APPLICANT: Mathew, Christopher George Porter  
TITLE OF INVENTION: Fanconi Anemia Type C Gene  
NUMBER OF SEQUENCES: 73  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Richard J. Polley, Esq.  
STREET: Klarquist, Sparkman, Campbell, Leigh &  
CITY: Portland  
STATE: Oregon  
COUNTRY: U.S.A.  
ZIP: 97204  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Disk, 3+1/2-inch  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: MS DOS  
SOFTWARE: Wordperfect 5.1/ASCII Text File  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/441, 430  
FILING DATE: May 15, 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. 07/876, 285  
FILING DATE: April 29, 1992  
APPLICATION NUMBER: U.S. 07/918, 313  
FILING DATE: July 21, 1992  
APPLICATION NUMBER: U.S. 08/003, 963  
FILING DATE: January 15, 1993



```

;
; ATTORNEY/AGENT INFORMATION:
; NAME: Clough, David W
; REGISTRATION NUMBER: 35,107
; REFERENCE/DOCKET NUMBER: 28111/32372
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; INFORMATION FOR SEQ ID NO: 233:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 39 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-350-260A-233

Alignment Scores:
Pred. No.: 357 Length: 39
Score: 45.00 Matches: 8
Percent Similarity: 66.67% Conservative: 0
Best Local Similarity: 66.67% Mismatches: 4
Query Match: 10.18% Indels: 0
DB: 2 Gaps: 0

US-09-599-087b-5 (1-81) x US-08-350-260A-233 (1-39)

QY 32 TTPserGlyArgArgThrArgLeuGlyCysHisArg 43
Db 3 TGTGCTCTCCGCGAATACCAAGCTGCTGTCATAGG 38

RESULT 6
US-08-642-255-23
; Sequence 23, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/B1R
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "synthetic"
US-08-642-255-23

Alignment Scores:
Pred. No.: 516 Length: 42
Score: 44.00 Matches: 7
Percent Similarity: 75.00% Conservative: 2
Best Local Similarity: 58.33% Mismatches: 3
Query Match: 9.95% Indels: 0
DB: 1 Gaps: 0

US-09-599-087b-5 (1-81) x US-08-642-255-23 (1-42)

QY 27 ArpProAlaLysAlaLysArgLysArgArgThrArg 38
Db 4 AGGCCAGAGAGCTGCTGCGGTCCAGCGCGACACAG 39

RESULT 7
US-07-903-456-7/C
; Sequence 7, Application us/07903456
; Patent No. 5574144
; GENERAL INFORMATION:
; APPLICANT: KAMBOJ, Rajender
; APPLICANT: ELLIOTT, Candace
; APPLICANT: NUTT, Stephen
; TITLE OF INVENTION: KAINATE-BINDING HUMAN CNS RECEPTORS OF
; TITLE OF INVENTION: THE EA4 FAMILY
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/903,456
; FILING DATE: 19920624
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 16777/183/ALLE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 889149
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-07-903-456-7

Alignment Scores:
Pred. No.: 657 Length: 50
Score: 44.00 Matches: 7
Percent Similarity: 60.00% Conservative: 2
Best Local Similarity: 46.67% Mismatches: 6
Query Match: 9.95% Indels: 0
DB: 1 Gaps: 0

US-09-599-087b-5 (1-81) x US-07-903-456-7 (1-50)

QY 34 GLyArgArgThrArgLeuGlyCysHisArgValProSerProAsn 48
Db 46 GGCATGAGCTCAGAACCTGCTGTCATGAGAGCTCAACTCCAAAC 2

RESULT 8
US-07-903-456-8/C
; Sequence 8, Application US/07903456
```



TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-249-241-8

Alignment Scores:  
Pred. No.: 657 Length: 50  
Score: 44.00 Matches: 7  
Percent Similarity: 60.00% Conservative: 2  
Best Local Similarity: 46.67% Mismatches: 6  
Query Match: 9.95% Indels: 0  
DB: 4 Gaps: 0

US-09-599-087b-5 (1-81) x US-08-249-241-8 (1-50)

Qy 34 GtAgtAgtThArGLeucYcShsArGvaLProSerProAsn 48  
Db 46 GGCATGAGCTCGAAGCTGTGTGATGAGAGCTCCACTCCAAAC 2

RESULT 11  
US-08-350-260A-287  
Sequence 287, Application US/08350260A  
Patent No. 5962255  
GENERAL INFORMATION:  
APPLICANT: Winter, Gregory Paul  
APPLICANT: Griffiths, Andrew David  
APPLICANT: Williams, Samuel Cameron  
APPLICANT: Waterhouse, Peter  
APPLICANT: Nissim, Ahuva  
APPLICANT: Johnson, Kevin Stuart  
TITLE OF INVENTION: Methods for producing members of specific  
TITLE OF INVENTION: binding pairs  
NUMBER OF SEQUENCES: 602  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David W. Clough  
STREET: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/350,260A  
FILING DATE: 05-DEC-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9110549.4  
FILING DATE: 15-MAY-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9206318.9  
FILING DATE: 24-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB91/01134  
FILING DATE: 10-JUL-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB92/00883  
FILING DATE: 15-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB93/00605  
FILING DATE: 24-MAR-1993

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/150,002  
FILING DATE: 31-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/307,619  
FILING DATE: 16-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Clough, David W  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 28111/32372  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
INFORMATION FOR SEQ ID NO: 287:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 51 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-350-260A-287

Alignment Scores:  
Pred. No.: 676 Length: 51  
Score: 44.00 Matches: 9  
Percent Similarity: 62.50% Conservative: 1  
Best Local Similarity: 56.25% Mismatches: 4  
Query Match: 9.95% Indels: 2  
DB: 2 Gaps: 1

US-09-599-087b-5 (1-81) x US-08-350-260A-287 (1-51)

Qy 32 TtpSerGtAgtAgtThArGLeucYs-----CyShsArGvaLPro, 45  
Db 3 TGTGCTCCGCGCCGAGATAGAGAGTGTGCTCCACTGCAATGAGTGTGCC 50

RESULT 12  
US-08-350-260A-290  
Sequence 290, Application US/08350260A  
Patent No. 5962255  
GENERAL INFORMATION:  
APPLICANT: Winter, Gregory Paul  
APPLICANT: Griffiths, Andrew David  
APPLICANT: Williams, Samuel Cameron  
APPLICANT: Waterhouse, Peter  
APPLICANT: Nissim, Ahuva  
APPLICANT: Johnson, Kevin Stuart  
TITLE OF INVENTION: Methods for producing members of specific  
TITLE OF INVENTION: binding pairs  
NUMBER OF SEQUENCES: 602  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David W. Clough  
STREET: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/350,260A  
FILING DATE: 05-DEC-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9110549.4  
FILING DATE: 15-MAY-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9206318.9  
FILING DATE: 24-MAR-1992  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/GB91/01134  
FILING DATE: 10-JUL-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB92/00883  
FILING DATE: 15-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB93/00605  
FILING DATE: 24-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/150,002  
FILING DATE: 31-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/307,619  
FILING DATE: 16-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: CLOUGH, DAVID W  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 28111/32372  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
INFORMATION FOR SEQ ID NO: 290:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 51 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-350-260A-290

Alignment Scores:  
Pred. No.: 676  
Score: 44.00  
Percent Similarity: 61.11%  
Best Local Similarity: 50.00%  
Query Match: 9.95%  
DB: 2  
Gaps: 1

US-09-599-087B-5 (1-81) x US-08-350-260A-290 (1-51)  
QY 32 TTPSERClyArGArGThArGLeucyCySHsArGValProSerProAsnSer 49  
Db 3 TGTCCCTCGCGCGGATACNMNMGTCGTGC-----TGTCCACGCGTCGACAGT 50

RESULT 13  
346668-7  
Patent No. 5466668  
APPLICANT: GLASER, CHARLES B.; MORSER, MICHAEL J.; LIGHT,  
DAVID R.  
TITLE OF INVENTION: SUPERIOR THROMBOMODULIN ANALOGS FOR  
PHARMACEUTICAL USE  
NUMBER OF SEQUENCES: 57  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/155,346  
FILING DATE: 22-NOV-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 568,456  
FILING DATE: 15-AUG-1990  
APPLICATION NUMBER: 506,325  
FILING DATE: 09-APR-1990  
APPLICATION NUMBER: 406,941  
FILING DATE: 13-SEP-1989  
APPLICATION NUMBER: 345,374  
FILING DATE: 28-APR-1989  
SEQ ID NO: 7:  
LENGTH: 72  
546668-7

Alignment Scores:  
Pred. No.: 1,42e+03  
Score: 43.00  
Percent Similarity: 83.33%  
Best Local Similarity: 58.33%  
Query Match: 9.73%  
DB: 6  
Gaps: 0

US-09-599-087B-5 (1-81) x 546668-7 (1-72)  
QY 9 LeuleucyslleleuleuleucySpheSerIlePhe 20  
Db 22 CTCTGCTGTGTGCTGCTGTGTGTGAGAGAGTCTTC 57

RESULT 14  
US-08-338-992B-5  
Sequence 5, Application US/08338992B  
Patent No. 5736524  
GENERAL INFORMATION:  
APPLICANT: CONTENT, JEAN  
APPLICANT: HUYGEN, KRIS  
APPLICANT: LIU, MARGARET A.  
APPLICANT: MONTGOMERY, DONNA  
APPLICANT: ULMER, JEFFREY  
TITLE OF INVENTION: A POLYNUCLEOTIDE TUBERCULOSIS VACCINE  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MICHAEL D. YABLONSKY - MERCK & CO., INC.  
STREET: 126 E. LINCOLN AVE., - P.O. BOX 2000  
CITY: RAHWAY  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07065-0900

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/338,992B  
FILING DATE: 14-NOV-1994  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: YABLONSKY, MICHAEL D.  
REGISTRATION NUMBER: 40,407  
REFERENCE/DOCKET NUMBER: 19342  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 908-594-4678  
TELEFAX: 908-594-4720

INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-338-992B-5

Alignment Scores:  
Pred. No.: 1,59e+03  
Score: 43.00  
Percent Similarity: 83.33%  
Best Local Similarity: 58.33%  
Query Match: 9.73%  
DB: 1  
Gaps: 0

US-09-599-087B-5 (1-81) x US-08-338-992B-5 (1-78)  
QY 9 LeuleucyslleleuleuleucySpheSerIlePhe 20  
Db 29 CTCTGCTGTGTGCTGCTGTGTGTGAGAGAGTCTTC 64

RESULT 15  
US-08-338-992B-6/C  
Sequence 6, Application US/08338992B  
Patent No. 5736524  
GENERAL INFORMATION:  
APPLICANT: CONTENT, JEAN  
APPLICANT: HUYGEN, KRIS  
APPLICANT: LIU, MARGARET A.



APPLICANT: MONTGOMERY, DONNA  
APPLICANT: GIMER, JEFFREY  
TITLE OF INVENTION: A POLYNUCLEOTIDE TUBERCULOSIS VACCINE  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MICHAEL D. YABLONSKY - MERCK & CO., INC.  
STREET: 126 E. LINCOLN AVE., - P.O. BOX 2000  
CITY: RAHWAY  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07065-0900  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/338,992B  
FILING DATE: 14-NOV-1994  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: YABLONSKY, MICHAEL D.  
REGISTRATION NUMBER: 40,407  
REFERENCE/DOCKET NUMBER: 19342  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 908-594,4678  
TELEFAX: 908-594-4720  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
OS-08-338-992B-6

Alignment Scores:	
Pred. No.:	1.59e+03
Score:	43.00
Percent Similarity:	83.33%
Best Local Similarity:	58.33%
Query Match:	9.73%
DB:	1
	Gaps: 0
	Indels: 0
	Mismatches: 2
	Conservative: 3
	Matches: 7
	Length: 78

US-09-599-087B-5 (1-81) x US-08-338-992B-6 (1-78)

QY 9 LeuLeuCySILLeuLeuLeuLeuCySPhseSerIlephe 20  
 ||| |||:::||||||| ||| :::::||||  
 Db 54 CTCTGCTGTGTGCTGCTGCTGTGTGTGAGACAGTCTTC 199

Search completed: May 8, 2003, 22:37:37  
Job time : 63 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: May 8, 2003, 21:49:18 ; Search time 103 Seconds  
(without alignments)  
977.273 Million cell updates/sec

Title: US-09-599-087B-5

Perfect score: 442  
Sequence: 1 MRLVLSLLCILLCFISF.....PCKLEPRRLVWVGPALPOV 81

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5	
Fgapop 6.0	Fgapext 7.0	
Delop 6.0	Delext 7.0	

Searched: 783854 segs, 621352466 residues

Total number of hits satisfying chosen parameters: 238982

Minimum DB seq length: 25  
Maximum DB seq length: 80

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+pn.model -DEV=xlp  
-O=/cg2\_1/USPTO.spool/US09599087/runat.07052003.113322.7398/app.query.fasta\_1.263  
-DB=Published\_Applications\_NA -QMT=fastap -SUFFIX=rnpb -MIMATCH=0.1  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bl0sum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR=SCORE-pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500  
-MLEN=25 -MAXLEN=80 -USER=US09599087 @CGN.1.1.93 @runat.07052003.113322.7398  
-NCPU=6 -ICPU=3 -NO\_ILPX -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT  
-SEPBLOCK=100 -LONGLOG -DEV\_TIMEOUT=120 -MARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10  
-XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published\_Applications\_NA\*

1:	/cg2_6/ptodata/1/pubpna/US07_PUBCOMB.seq*
2:	/cg2_6/ptodata/1/pubpna/PTC_NEW_PUB.seq*
3:	/cg2_6/ptodata/1/pubpna/US06_NEW_PUB.seq*
4:	/cg2_6/ptodata/1/pubpna/US06_PUBCOMB.seq*
5:	/cg2_6/ptodata/1/pubpna/US07_NEW_PUB.seq*
6:	/cg2_6/ptodata/1/pubpna/PTUB_PUBCOMB.seq*
7:	/cg2_6/ptodata/1/pubpna/US08_NEW_PUB.seq*
8:	/cg2_6/ptodata/1/pubpna/US08_PUBCOMB.seq*
9:	/cg2_6/ptodata/1/pubpna/US09_NEW_PUB.seq*
10:	/cg2_6/ptodata/1/pubpna/US09_PUBCOMB.seq*
11:	/cg2_6/ptodata/1/pubpna/US10_NEW_PUB.seq*
12:	/cg2_6/ptodata/1/pubpna/US10_PUBCOMB.seq*
13:	/cg2_6/ptodata/1/pubpna/US60_NEW_PUB.seq*
14:	/cg2_6/ptodata/1/pubpna/US60_PUBCOMB.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	44	10.0	50	US-10-126-617-7	Sequence 7, Appl1
c 2	44	10.0	50	US-10-126-617-8	Sequence 8, Appl1
c 3	43	9.7	69	US-10-159-428-1	Sequence 1, Appl1
c 4	43	9.7	78	US-10-204-200-6	Sequence 6, Appl1

Result No.	Score	Query Match	Length	ID	Description
c 5	43	9.7	78	US-10-204-200-7	Sequence 7, Appl1
c 6	43	9.7	78	US-09-950-844-5	Sequence 5, Appl1
c 7	43	9.7	78	US-09-950-844-6	Sequence 6, Appl1
c 8	43	9.7	78	US-09-835-699-5	Sequence 5, Appl1
c 9	43	9.7	78	US-09-835-699-6	Sequence 6, Appl1
c 10	42.5	9.6	59	US-10-006-593-89	Sequence 89, Appl1
c 11	42	9.5	50	US-10-137-788-6	Sequence 6, Appl1
c 12	41.5	9.4	46	US-09-827-864-19	Sequence 19, Appl1
c 13	41.5	9.4	54	US-09-294-093B-4383	Sequence 4383, Ap
c 14	41.5	9.4	60	US-09-902-941-1979	Sequence 1979, Ap
c 15	41.5	9.4	60	US-10-017-754-1979	Sequence 1979, Ap
c 16	41	9.3	60	US-09-954-692-13	Sequence 13, Appl1
c 17	41	9.3	60	US-09-559-671A-13	Sequence 13, Appl1
c 18	41	9.3	80	US-09-440-820-17	Sequence 17, Appl1
c 19	41	9.3	80	US-09-157-748-43	Sequence 43, Appl1
c 20	40.5	9.2	75	US-09-864-761-28813	Sequence 28813, A
c 21	40	9.0	31	US-09-773-599-16	Sequence 16, Appl1
c 22	40	9.0	31	US-09-801-274-1511	Sequence 1511, Ap
c 23	40	9.0	31	US-09-801-274-1512	Sequence 1512, Ap
c 24	40	9.0	46	US-09-940-244-227	Sequence 227, App
c 25	40	9.0	46	US-09-877-478-5348	Sequence 5348, Ap
c 26	40	9.0	59	US-10-025-380-976	Sequence 976, App
c 27	40	9.0	59	US-09-922-217-976	Sequence 976, App
c 28	40	9.0	59	US-09-833-263-976	Sequence 976, App
c 29	40	9.0	59	US-10-013-737-15	Sequence 15, Appl1
c 30	40	9.0	60	US-09-902-941-1880	Sequence 1980, Ap
c 31	40	9.0	60	US-10-017-754-1980	Sequence 1980, Ap
c 32	40	9.0	63	US-09-923-876-1639	Sequence 1639, Ap
c 33	40	9.0	64	US-10-072-036-7	Sequence 7, Appl1
c 34	40	9.0	75	US-09-979-999-15	Sequence 15, Appl1
c 35	40	9.0	75	US-10-045-465-2	Sequence 2, Appl1
c 36	40	9.0	78	US-09-864-761-21687	Sequence 21687, A
c 37	39.5	8.9	78	US-09-861-257-47	Sequence 47, Appl1
c 38	39	8.8	25	US-10-215-112-13778	Sequence 13778, A
c 39	39	8.8	25	US-10-156-634A-13	Sequence 13, Appl1
c 40	39	8.8	36	US-09-345-373-5	Sequence 5, Appl1
c 41	39	8.8	37	US-09-263-959-815	Sequence 815, App
c 42	39	8.8	40	US-09-263-959-458	Sequence 458, App
c 43	39	8.8	64	US-10-085-906-168	Sequence 168, App
c 44	39	8.8	65	US-09-747-003-12	Sequence 12, Appl1
c 45	39	8.8	69	US-10-076-816-28	Sequence 28, Appl1

## ALIGNMENTS

RESULT 1  
US-10-126-617-7/c  
Sequence 7, Application US/10126617  
Patent No. US20020168723A1  
GENERAL INFORMATION:  
APPLICANT: KAMBOJ, RAJENDER  
APPLICANT: ELLIOTT, CANDACE E.  
APPLICANT: NUTT, STEPHEN L.  
TITLE OF INVENTION: KINASE-BINDING HUMAN CNS RECEPTORS OF THE EAA FAMILY  
FILE REFERENCE: 016777/0471  
CURRENT APPLICATION NUMBER: US/10/126,617  
CURRENT FILING DATE: 2002-04-22  
PRIOR APPLICATION NUMBER: 08/249,241  
PRIOR FILING DATE: 1994-05-25  
PRIOR APPLICATION NUMBER: 07/903,456  
PRIOR FILING DATE: 1992-06-24  
NUMBER OF SEQ ID NOS: 9  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 7  
LENGTH: 50  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-126-617-7

Alignment Scores:  
Pred. No.: 566  
Score: 44.00  
Percent Similarity: 60.00%  
Length: 50  
Matches: 7  
Conservative: 2

; PRIOR FILING DATE: 2000-02-17

; PRIOR FILING DATE: 2000-02-17

```

; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Human
US-10-204-200-7

Alignment Scores:
Pred. No.: 1.32e+03 Length: 78
Score: 43.00 Matches: 7
Percent Similarity: 83.33% Conservative: 3
Best Local Similarity: 58.33% Mismatches: 2
Query Match: 9.73% Indels: 0
DB: Gaps: 0

US-09-599-087b-5 (1-81) x US-10-204-200-7 (1-78)
QY 9 Leuleucysilleuleucyspheserillephe 20
||| |||:||||| |||: |||: |||
54 CTCTGCTGTGCTGCTGCTGTGTGAGCAGCTCTTC 19

RESULT 6
US-09-950-844-5
; Sequence 5, Application US/09950844
; Patent No. US20020045594A1
; GENERAL INFORMATION:
; APPLICANT: Volklin, David B.
; APPLICANT: Evans, Robert K.
; APPLICANT: Ulmer, Jeffrey B.
; APPLICANT: Caulfield, Michael J.
; TITLE OF INVENTION: POLYNUCLEOTIDE VACCINE FORMULATIONS
; FILE REFERENCE: 19907YIACA
; CURRENT APPLICATION NUMBER: US/09/950,844
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 09/112,655
; PRIOR FILING DATE: 1998-07-09
; PRIOR APPLICATION NUMBER: 09/023,834
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: 60/038,194
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Oligonucleotide
US-09-950-844-5

Alignment Scores:
Pred. No.: 1.32e+03 Length: 78
Score: 43.00 Matches: 7
Percent Similarity: 83.33% Conservative: 3
Best Local Similarity: 58.33% Mismatches: 2
Query Match: 9.73% Indels: 0
DB: Gaps: 0

US-09-599-087b-5 (1-81) x US-09-950-844-5 (1-78)
QY 9 Leuleucysilleuleucyspheserillephe 20
||| |||:||||| |||: |||: |||
29 CTCTGCTGTGCTGCTGCTGTGTGAGCAGCTCTTC 64

RESULT 7
US-09-950-844-6/c
; Sequence 6, Application US/09950844
; Patent No. US20020045594A1
; GENERAL INFORMATION:
; APPLICANT: Volklin, David B.
; APPLICANT: Evans, Robert K.
; APPLICANT: Ulmer, Jeffrey B.
; APPLICANT: Caulfield, Michael J.
; TITLE OF INVENTION: POLYNUCLEOTIDE VACCINE FORMULATIONS
; FILE REFERENCE: 19907YIACA
; CURRENT APPLICATION NUMBER: US/09/950,844
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 09/112,655
; PRIOR FILING DATE: 1998-07-09
; PRIOR APPLICATION NUMBER: 09/023,834
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: 60/038,194
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Oligonucleotide
US-09-950-844-6

Alignment Scores:
Pred. No.: 1.32e+03 Length: 78
Score: 43.00 Matches: 7
Percent Similarity: 83.33% Conservative: 3
Best Local Similarity: 58.33% Mismatches: 2
Query Match: 9.73% Indels: 0
DB: Gaps: 0

US-09-599-087b-5 (1-81) x US-09-950-844-6 (1-78)
QY 9 Leuleucysilleuleucyspheserillephe 20
||| |||:||||| |||: |||: |||
54 CTCTGCTGTGCTGCTGCTGTGTGAGCAGCTCTTC 19

RESULT 8
US-09-835-699-5
; Sequence 5, Application US/09835699
; Patent No. US20020147167A1
; GENERAL INFORMATION:
; APPLICANT: Armstrong, Marcy E.
; APPLICANT: Keys, Robert D.
; APPLICANT: Lewis, John A.
; APPLICANT: Liu, Margaret A.
; APPLICANT: McClements, William L.
; TITLE OF INVENTION: A POLYNUCLEOTIDE HERPES VIRUS VACCINE
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: P.O. Box 2000, 126 E. Lincoln Ave.
; CITY: Rahway
; STATE: NJ
; COUNTRY: USA
; ZIP: 07065-0907
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/835,699
; FILING DATE: 17-Aug-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PC9/US97/17262
; FILING DATE: 26-SEP-1997
; APPLICATION NUMBER: US 08/720,758
; FILING DATE: 01-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hand, J. Mark
; REGISTRATION NUMBER: 36,545
; REFERENCE/DOCKET NUMBER: 19238CC
```

```

TELECOMMUNICATION INFORMATION:
  TELEPHONE: 732-594-3905
  TELEFAX: 732-594-4720
  TELEX: <Unknown>
  INFORMATION FOR SEQ ID NO: 5:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 78 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  MOLECULE TYPE: Other
  FEATURE:
    NAME/KEY:
    LOCATION: 1..78
    OTHER INFORMATION: nucleic acid
  SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-835-699-5

Alignment Scores:
  Pred. No.: 1.32e+03 Length: 78
  Score: 43.00 Matches: 7
  Percent Similarity: 83.33% Conservative: 3
  Best Local Similarity: 58.33% Mismatches: 2
  Query Match: 9.73% Indels: 0
  DB: 10 Gaps: 0

US-09-599-087b-5 (1-81) x US-09-835-699-5 (1-78)
Qy 9 LeuLeuCySileLeuLeuLeuCySPheserIlePhe 20
Db 29 CTGCTGTGTGCTGCTGCTGTGTGAGACAGCTCTTC 64

RESULT 9
  US-09-835-699-6/c
  Sequence 6, Application US/09835699
  Patent NO. US20020147167A1
  GENERAL INFORMATION:
    APPLICANT: Armstrong, Marcy E.
    Keys, Robert D.
    Lewis, John A.
    Liu, Margaret A.
    McClements, William L.
  TITLE OF INVENTION: A POLYNUCLEOTIDE HERPES VIRUS
  VACCINE
  NUMBER OF SEQUENCES: 15
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Merck & Co., Inc.
    STREET: P.O. Box 2000, 126 E. Lincoln Ave.
    CITY: Rahway
    STATE: NJ
    COUNTRY: USA
    ZIP: 07065-0907
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Diskette
    COMPUTER: IBM Compatible
    OPERATING SYSTEM: Windows
    SOFTWARE: FastSeq for Windows Version 2.0b
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/09/835.699
    FILING DATE: 17-Aug-2001
    CLASSIFICATION: <Unknown>
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: PCT/US97/17262
    FILING DATE: 26-SEP-1997
    APPLICATION NUMBER: US 08/720,758
    FILING DATE: 01-OCT-1996
  ATTORNEY/AGENT INFORMATION:
    NAME: Hand, J. Mark
    REGISTRATION NUMBER: 36,545
    REFERENCE/DOCKET NUMBER: 19258CC
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 732-594-3905
    TELEFAX: 732-594-4720

```

```

TELEX: <Unknown>
  INFORMATION FOR SEQ ID NO: 6:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 78 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  MOLECULE TYPE: Other
  FEATURE:
    NAME/KEY:
    LOCATION: 1..78
    OTHER INFORMATION: nucleic acid
  SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-835-699-6

Alignment Scores:
  Pred. No.: 1.32e+03 Length: 78
  Score: 43.00 Matches: 7
  Percent Similarity: 83.33% Conservative: 3
  Best Local Similarity: 58.33% Mismatches: 2
  Query Match: 9.73% Indels: 0
  DB: 10 Gaps: 0

US-09-599-087b-5 (1-81) x US-09-835-699-6 (1-78)
Qy 9 LeuLeuCySileLeuLeuLeuCySPheserIlePhe 20
Db 54 CTGCTGTGTGCTGCTGCTGTGTGAGACAGCTCTTC 19

RESULT 10
  US-10-006-593-89/c
  Sequence 89, Application US/10006593
  Publication No. US20030049683A1
  GENERAL INFORMATION:
    APPLICANT: Bowdish, Katherine S.
    APPLICANT: Frederickson, Shana
    APPLICANT: Renshaw, Mark
  TITLE OF INVENTION: RATIONALLY DESIGNED ANTIBODIES
  FILE REFERENCE: 1087-2
  CURRENT APPLICATION NUMBER: US/10/006.593
  CURRENT FILING DATE: 2001-12-05
  PRIOR APPLICATION NUMBER: US 60/251,448
  PRIOR FILING DATE: 2000-12-05
  PRIOR APPLICATION NUMBER: US 60/288,889
  PRIOR FILING DATE: 2001-05-04
  PRIOR APPLICATION NUMBER: US 60/294,068
  PRIOR FILING DATE: 2001-05-29
  NUMBER OF SEQ ID NOS: 118
  SOFTWARE: PatentIn version 3.1
  SEQ ID NO 89
  LENGTH: 59
  TYPE: DNA
  ORGANISM: artificial sequence
  FEATURE:
    OTHER INFORMATION: primer
  US-10-006-593-89

Alignment Scores:
  Pred. No.: 1.07e+03 Length: 59
  Score: 42.50 Matches: 9
  Percent Similarity: 57.89% Conservative: 2
  Best Local Similarity: 47.37% Mismatches: 7
  Query Match: 9.62% Indels: 1
  DB: 9 Gaps: 1

US-09-599-087b-5 (1-81) x US-10-006-593-89 (1-59)
Qy 15 LeuCySPheserIlePheSerThrGluGlyLysArgArgPro---AlaLysAlaTrp 32
Db 59 CTGCTATTTCATTATTCATAAACCAGATCGAAGGCCCAACCCCTGCGCAGTGGCTGG 3

RESULT 11
US-10-137-788-6

```

Sequence 6, Application US/10137788  
Publication No. US20030039636A1  
GENERAL INFORMATION:  
APPLICANT: Leboulch, Philippe  
TITLE OF INVENTION: NOVEL SELF-INACTIVATING (SIN) LENTIVIRAL VECTORS  
FILE REFERENCE: ICI-023  
CURRENT APPLICATION NUMBER: US/10/137,788  
CURRENT FILING DATE: 2002-05-01  
PRIOR APPLICATION NUMBER: US 60/288,042  
PRIOR FILING DATE: 2001-05-01  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: fastseq for Windows Version 4.0  
SEQ ID NO 6  
LENGTH: 50  
TYPE: DNA  
ORGANISM: Human Immunodeficiency virus type 1  
US-10-137-788-6

Alignment Scores:  
Pred. No.: 1.01e+03 Length: 50  
Score: 42.00 Matches: 7  
Percent Similarity: 73.33% Conservative: 4  
Best Local Similarity: 46.67% Mismatches: 4  
Query Match: 9.50% Indels: 0  
DB: Gaps: 0

US-09-599-087b-5 (1-81) x US-10-137-788-6 (1-50)

OY 4 LeuValLeuSerLeuLeuCysAlaLeuLeuLeuCysPheSer 18  
Db 4 CTGCGCTGAGTGCCTCAATGCTGTGTTTGTGTCTG 48

RESULT 12  
US-09-827-864-19/c  
Sequence 19, Application US/09827864  
Patent No. US20020009458A1  
GENERAL INFORMATION:  
APPLICANT: COLAU, DIDIER  
ROOS, JOEL  
TITLE OF INVENTION: RECOMBINANT DNA SEQUENCES, MOLECULES,  
VECTORS AND VACCINES FOR FELINE  
CALICIVIRUS DISEASE AND METHODS FOR  
PRODUCING AND USING SAME  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MCDELMOTT, WILL & EMERY  
STREET: 1850 K STREET, N.W., SUITE 500  
CITY: WASHINGTON  
STATE: DC  
COUNTRY: USA  
ZIP: 20006  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/827,864  
FILING DATE: 06-Apr-2001  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: GADIANO, WILHEM F  
REGISTRATION NUMBER: 37,136  
REFERENCE/DOCKET NUMBER: 37712-213  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 778-8373  
TELEFAX: (202) 778-8335  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 46 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: nucleic acid  
SEQUENCE DESCRIPTION: SEQ ID NO: 19:  
US-09-827-864-19

Alignment Scores:  
Pred. No.: 1.05e+03 Length: 46  
Score: 41.50 Matches: 8  
Percent Similarity: 64.29% Conservative: 1  
Best Local Similarity: 57.14% Mismatches: 4  
Query Match: 9.39% Indels: 1  
DB: Gaps: 1

US-09-599-087b-5 (1-81) x US-09-827-864-19 (1-46)

OY 60 CysLysProCysLysLeuGluProGluProArgLeuTrpVal 73  
Db 42 TGTCAACCTTGCAAGCTGGAATGACGCTCTCT--TGCTTA 4

RESULT 13  
US-09-294-093B-4383/c  
Sequence 4383, Application US/09294093B  
Patent No. US20010051335A1  
GENERAL INFORMATION:  
APPLICANT: Lalugudi, Raghunath, V.  
APPLICANT: Ito, Laura, Y.  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL  
FILE REFERENCE: PL-0009 US  
CURRENT APPLICATION NUMBER: US/09/294,093B  
CURRENT FILING DATE: 1999-04-16  
PRIOR APPLICATION NUMBER: 60/082,567  
PRIOR FILING DATE: April 21, 1998  
NUMBER OF SEQ ID NOS: 6207  
SOFTWARE: PERL Program  
SEQ ID NO 4383  
LENGTH: 54  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
NAME/KEY: misc.feature  
OTHER INFORMATION: Incyte ID NO. US20010051335A1 700353949H1  
NAME/KEY: unsure  
LOCATION: 34  
OTHER INFORMATION: a, t, c, g, or other  
US-09-294-093B-4383

Alignment Scores:  
Pred. No.: 1.28e+03 Length: 54  
Score: 41.50 Matches: 8  
Percent Similarity: 69.23% Conservative: 1  
Best Local Similarity: 61.54% Mismatches: 1  
Query Match: 9.39% Indels: 3  
DB: Gaps: 1

US-09-599-087b-5 (1-81) x US-09-294-093B-4383 (1-54)

OY 34 GlyArgArgThrArgLeuCysCysHisArgValProSer 46  
Db 52 GGACGGAGG-----TGCTGCATNTATATACCAAGT 23

RESULT 14  
US-09-902-941-1979  
Sequence 1979, Application US/09902941  
Patent No. US20020172952A1  
GENERAL INFORMATION:  
APPLICANT: Henderson, Robert A.  
APPLICANT: Wang, Tongtong  
APPLICANT: Watanabe, Yoshihiro  
APPLICANT: Johnson, Jeffrey C.  
APPLICANT: Retter, Marc W.  
APPLICANT: Marnerakis, Margarita  
APPLICANT: Carter, Darrick

Search completed: May 8, 2003, 22:39:34  
Job time : 104 secs

APPLICANT: Fanger, Gary R.  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Bangur, Chaitanya S.  
APPLICANT: McNabb, Andia  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
FILE REFERENCE: 210121.478C17  
CURRENT APPLICATION NUMBER: US/09/902,941  
CURRENT FILING DATE: 2001-07-10  
NUMBER OF SEQ ID NOS: 2002  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 1979  
LENGTH: 60  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-902-941-1979

Alignment Scores:  
Pred. No.: 1.46e+03 Length: 60  
Score: 41.50 Matches: 9  
Percent Similarity: 61.11% Conservative: 2  
Best Local Similarity: 50.00% Mismatches: 6  
Query Match: 9.39% Indels: 1  
DB: 9 Gaps: 1

US-09-599-087b-5 (1-81) x US-09-902-941-1979 (1-60)

Qy 34 G1yArygArGThrArgLeu--CysCysHsArgValProSerProAsnSerThr 50  
||| : : : ||| ||||| ||||| ||||| |||||  
Db 1 GGGGGCTCAGCGAGCTGGGGTGTGTGTGGGGGTATCCGAGTCCCGAAGACACC 54

RESULT 15  
US-10-017-754-1979

Sequence 1979, Application US/10017754  
Publication No. US20030054363A1  
GENERAL INFORMATION:  
APPLICANT: Henderson, Robert A.  
APPLICANT: Wang, Tonglong  
APPLICANT: Watanabe, Yoshihiro  
APPLICANT: Johnson, Jeffrey C.  
APPLICANT: Retter, Marc W.  
APPLICANT: Marnierakis, Margarita  
APPLICANT: Carter, Darrick  
APPLICANT: Fanger, Gary R.  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Bangur, Chaitanya S.  
APPLICANT: McNabb, Andia  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
FILE REFERENCE: 210121.478C18  
CURRENT APPLICATION NUMBER: US/10/017,754  
CURRENT FILING DATE: 2001-10-29  
NUMBER OF SEQ ID NOS: 2004  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 1979  
LENGTH: 60  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-017-754-1979

Alignment Scores:  
Pred. No.: 1.46e+03 Length: 60  
Score: 41.50 Matches: 9  
Percent Similarity: 61.11% Conservative: 2  
Best Local Similarity: 50.00% Mismatches: 6  
Query Match: 9.39% Indels: 1  
DB: 9 Gaps: 1

US-09-599-087b-5 (1-81) x US-10-017-754-1979 (1-60)

Qy 34 G1yArygArGThrArgLeu--CysCysHsArgValProSerProAsnSerThr 50  
||| : : : ||| ||||| ||||| ||||| |||||  
Db 1 GGGGGCTCAGCGAGCTGGGGTGTGTGTGGGGGTATCCGAGTCCCGAAGACACC 54

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 21:41:38 ; Search time 35 Seconds  
(without alignments)  
308.380 Million cell updates/sec

Title: US-09-599-087B-5

Perfect score: 442

Sequence: 1 MRLVLSILLCITLLCFSEIF.....PKLEPEPRLMWPGALPOV 81

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 219634

Minimum DB seq length: 25  
Maximum DB seq length: 80

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A.GeneSeq\_101002:\*

- 1: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT:\*
- 2: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT:\*
- 3: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT:\*
- 4: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT:\*
- 5: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT:\*
- 6: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT:\*
- 7: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT:\*
- 8: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT:\*
- 9: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT:\*
- 10: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT:\*
- 11: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT:\*
- 12: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT:\*
- 13: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT:\*
- 14: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT:\*
- 15: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT:\*
- 16: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT:\*
- 17: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT:\*
- 18: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT:\*
- 19: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT:\*
- 20: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT:\*
- 21: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT:\*
- 22: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT:\*
- 23: /SIDS2/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	326	73.8	57	23	AAE16482
2	228.5	51.7	78	23	AAE16479
3	183	41.4	77	23	AAE16483
4	169	38.2	30	23	AAE16485
5	163	36.9	28	23	AAE16486
6	132.5	30.0	54	23	AAE16480
7	62	14.0	78	22	AAU45851
8	61	13.8	75	22	AAU65859
9	61	13.8	75	22	AAU54516
10	59.5	13.5	73	21	AAU00379

11	58.5	13.2	77	20	AAV36266	Human secreted pro
12	58	13.1	53	19	AAW75178	Human secreted pro
13	58	13.1	62	22	AAO05651	Human polypeptide
14	58	13.1	63	21	AAE44937	Human secreted pro
15	58	13.1	63	22	AAE60725	Human secreted pro
16	57.5	13.0	74	21	AAE12200	Arabidopsis thaliana
17	57.5	13.0	76	21	AAE12199	Arabidopsis thaliana
18	57	12.9	78	22	ABG08705	Novel human diago
19	56	12.7	58	22	AAE60722	Human secreted pro
20	56	12.7	61	22	AAO12472	Human polypeptide
21	56	12.7	65	22	AAU43919	Human secreted pro
22	55.5	12.6	53	20	AAV27661	Human secreted pro
23	55.5	12.6	64	22	AAU40623	Human secreted pro
24	55	12.4	50	22	AAU61037	Human secreted pro
25	55	12.4	50	22	AAU64737	Human secreted pro
26	55	12.4	80	20	AAV3650	Human secreted pro
27	54.5	12.3	76	22	AAU48701	Human secreted pro
28	54	12.2	57	22	AAU62049	Human secreted pro
29	54	12.2	57	22	AAU62049	Human secreted pro
30	54	12.2	68	22	AAU40407	Human secreted pro
31	54	12.2	70	21	AAE63096	Human secreted pro
32	53.5	12.1	44	22	AAE43990	Human secreted pro
33	53.5	12.1	44	22	AAE43990	Human secreted pro
34	53.5	12.1	44	22	AAE43990	Human secreted pro
35	53.5	12.1	44	22	AAE21629	Human secreted pro
36	53.5	12.1	44	22	AAE37931	Human secreted pro
37	53.5	12.1	52	22	AAE64672	Human secreted pro
38	53.5	12.1	53	22	AAE25839	Human secreted pro
39	53.5	12.1	62	22	AAE85736	Human secreted pro
40	53.5	12.1	71	22	AAU60257	Human secreted pro
41	53.5	12.1	71	23	AAU07526	Human secreted pro
42	53.5	12.1	74	22	AAO12569	Human secreted pro
43	53	12.0	51	22	AAU65879	Human secreted pro
44	53	12.0	61	22	AAU33442	Human secreted pro
45	53	12.0	61	22	AAU54661	Human secreted pro

#### ALIGNMENTS

RESULT 1	AAE16482	standard: peptide; 57 AA.
ID	AAE16482	
AC	AAE16482	
XX		
DT	09-APR-2002 (first entry)	
XX		
DE	Human Secreted epithelial colon stromal-1 (Secs-1) protein fragment.	
XX		
KW	Secreted epithelial colon stromal-1; Secs-1; gene therapy; osteopetrosis;	
KW	haematopoietic disorder; osteoporosis; osteogenesis imperfecta; cachexia;	
KW	Paget's disease; peridontal disease; hypercalcaemia; glomerulonephritis;	
KW	diabetes; obesity; osteopathic; cytostatic; nephrotoxic; antidiabetic;	
KW	anorectic; immunomodulator; antiproliferative; vulnery; antileukemia;	
KW	gynaecological; antineoplastic; antineoplastic; cancer; cell therapy;	
KW	human.	
OS	Homo sapiens.	
XX		
PN	WO200198497-A1.	
XX		
PD	27-DEC-2001.	
XX		
PR	28-NOV-2000; 2000WO-US32479.	
XX		
PR	21-JUN-2000; 2000US-0599087.	
XX		
PR	28-NOV-2000; 2000US-0724000.	
XX		
PA	(AMGE-) AMGEN INC.	
XX		
PI	Polverino AJ, Luethy R.	
XX		



DR	WPI; 2002-122281/16.
PX	Secreted epithelial colon stromal-1 polypeptides and nucleic acids,
PT	useful for diagnosing, treating and preventing hematopoietic disorder,
PT	osteoporosis, Paget's disease, cancer, diabetes -
XX	
PS	Claim 14; Page 122; 134pp; English.
CC	
CC	The present invention relates to an isolated murine or human secreted
CC	epithelial colon stromal-1 (Secs-1) polypeptide, its allelic or splice
CC	variant, orthologue, fragment or mutant. Secs-1 gene is used in gene
CC	therapy and cell therapy. Secs-1 is useful for identifying a compound
CC	which binds to a Secs-1 polypeptide. Secs-1 is useful for treating,
CC	preventing or ameliorating a disease condition such as haematopoietic
CC	disorder, osteoporosis, osteopetrosis, osteogenesis imperfecta, Paget's
CC	disease, periodontal disease, hypercalcaemia, acute glomerulonephritis,
CC	chronic glomerulonephritis, cancer, diabetes, obesity or cachexia. Secs-1
CC	is also useful for diagnosing a pathological condition which involves
CC	determining the presence or amount of Secs-1 or polypeptide encoded by
CC	Secs-1 DNA in a sample, and diagnosing a pathological condition, or
CC	susceptibility to pathological condition based on the presence or amount
CC	of expression of the polypeptide. The present sequence is human Secs-1
CC	protein fragment.
XX	
SQ	Sequence 57 AA:
Query Match	73.8%; Score 326; DB 23; Length 57;
Best Local Similarity	100.0%; Pred. No. 1e-30;
Matches 57; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	25 KRRPAKASGRTRCCHRVSPNSTNLKGHHVRLCKPCKLEPERLWVYGALPY 81       Db 1 KRRPAKASGRTRCCHRVSPNSTNLKGHHVRLCKPCKLEPERLWVYGALPY 57
RESULT 2	
AAE16479	AAE16479 standard; Protein; 78 AA.
XX	
AC	AAE16479;
XX	
DT	09-APR-2002 (first entry)
XX	
DE	Mouse Secreted epithelial colon stromal-1 (Secs-1) protein.
KM	Secreted epithelial colon stromal-1; Secs-1; gene therapy; osteopetrosis;
KM	hematopoietic disorder; osteoporosis; osteogenesis imperfecta; cachexia;
KM	Paget's disease; periodontal disease; hypercalcaemia; glomerulonephritis;
KM	diabetes; obesity; osteoarthritis; cytostatic; nephroretrophic; antidiabetic;
KM	anorectic; immunomodulator; antipsoriatic; vulnery; antiinfertility;
KM	gynaecological; antitumor; antinflammatory; cancer; cell therapy;
KM	mouse.
OS	Mus musculus.
XX	
FT	Key Location/Qualifiers
FT	Peptide 1..24
FT	/label= Signal_peptide
FT	Protein 25..78
FT	/label= Mature_mouse_Secs-1_protein
XX	
FN	MO200198497-A1.
XX	
PD	27-DEC-2001.
XX	
PF	28-NOV-2000; 2000WO-US32479.
XX	
PR	21-JUN-2000; 2000US-0599087.
XX	
PR	28-NOV-2000; 2000US-0724000.
XX	
PA	(AMGE-) AMGEN INC.
XX	
CI	Polverino AJ Luethy R;

	XX	DR	WPI: 2002-122281/16.
	XX	N-FSDB:	AAD27024.
	PT	Secreted epithelial colon stromal-1 polypeptides and nucleic acids,	
	PR	useful for diagnosing, treating and preventing hematopoietic disorder,	
	PT	osteoporosis; Paget's disease; cancer; diabetes -	
	PS	Claim 13; Fig 3; 134pp; English.	
	CC	The present invention relates to an isolated murine or human secreted	
	CC	epithelial colon stroma-1 (Secs-1) polypeptide, its allelic or splice	
	CC	variant, orthologue, fragment or mutant. Secs-1 gene is used in gene	
	CC	therapy and cell therapy. Secs-1 is useful for identifying a compound	
	CC	which binds to a Secs-1 polypeptide. Secs-1 is useful for treating,	
	CC	preventing or ameliorating a disease condition such as haematopoietic	
	CC	disorder, osteoporosis, osteopetrosis, osteogenesis imperfecta, Paget's	
	CC	disease, periodontal disease, hypercalcaemia, acute glomerulonephritis,	
	CC	chronic glomerulonephritis, cancer, diabetes, obesity or cachexia. Secs-1	
	CC	is also useful for diagnosing a pathological condition which involves	
	CC	determining the presence or amount of secs-1 or polypeptide encoded by	
	CC	Secs-1 DNA in a sample, and diagnosing a pathological condition, or	
	CC	susceptibility to pathological condition based on the presence or amount-	
	CC	of expression of the polypeptide. The present sequence is mouse Secs-1	
	CC	protein.	
SQ	Sequence	78 AA;	
OY	Query Match	Best Local Similarity    51.7%; Score 228.5; DB 23; Length 78; Matches     45; Conservative    55.6%; Pred. No. 3,Se=19; Matches      45; Conservative    10; Mismatches    23; Indels      3; Gaps      1	
Dy	1 MLRLVSLLCITLLCFISFTSGRRPKAWGRRTRLCGRVPSPNSTNLKHHVRIC 60          :             :             :             1 MRLLLSCGLWCILLFCIFSSSEGGRRHPAKSLIKLR--CCHLSPRSKLTTWKGHRTPC 57		
Oy	61 KPCKLEPPRILVVYPGALPOV 81 :::  :::  :::  : 58 RLCRNKLDPVKSWVPGALPOL 78		
ID	AAEL16483 standard; Protein; 77 AA. AAE16483		
AC	AAE16483;		
DT	09-APR-2002 (first entry)		
DE	Rat Secreted epithelial colon stromal-1 (Secs-1) protein.		
KM	Secreted epithelial colon stromal-1; Secs-1; gene therapy; osteopetrosis; haemato-oncologic disorders; osteoporosis; osteogenesis imperfecta; cachexia; Paget's disease; periodontal disease; hypercalcaemia; glomerulonephritis; diabetes; obesity; osteoarthritis; cytostatic; nephrotrophic; anti-diabetic; anorectic; immunomodulator; antiproliferative; vulnary; antifertility; gynaecological; antitumor; anti-inflammatory; cancer; cell therapy; rat.		
OS	Rattus norvegicus.		
PJ	M02001.98497-A1.		
PD	27-DEC-2001.		
PF	28-NOV-2000; 2000MO-US32479.		
PR	21-JUN-2000; 2000US-0599087.		
RA	(AMGE-) AMGEN INC.		
PI	Polverino AJ, Luethy R;		

XX WPI: 2002-122281/16.  
 DR Secreted epithelial colon stromal-1 polypeptides and nucleic acids,  
 XX useful for diagnosing, treating and preventing hematopoietic disorder,  
 PT osteoporosis, Paget's disease, cancer, diabetes -  
 XX Disclosure; Fig 3; 134pp; English.  
 PS  
 XX The present invention relates to an isolated murine or human secreted  
 CC epithelial colon stromal-1 (Secs-1) polypeptide, its allelic or splice  
 CC variant, orthologue, fragment or mutant. Secs-1 gene is used in gene  
 CC therapy and cell therapy. Secs-1 is useful for identifying a compound  
 CC which binds to a Secs-1 polypeptide. Secs-1 is useful for treating,  
 CC preventing or ameliorating a disease condition such as hematopoietic  
 CC disorder, osteoporosis, osteopetrosis, osteogenesis imperfecta, Paget's  
 CC disease, periodontal disease, hypercalcaemia, acute glomerulonephritis,  
 CC chronic glomerulonephritis, cancer, diabetes, obesity or cachexia. Secs-1  
 CC is also useful for diagnosing a pathological condition which involves  
 CC determining the presence or amount of Secs-1 or polypeptide encoded by  
 CC Secs-1 DNA in a sample; and diagnosing a pathological condition, or  
 CC susceptibility to pathological condition based on the presence or amount  
 CC of expression of the polypeptide. The present sequence is rat Secs-1  
 CC protein.  
 CC  
 XX Sequence 77 AA;  
 SQ  
 Query Match 41.4%; Score 183; DB 23; Length 77;  
 Best Local Similarity 46.9%; Pred. No. 7.2e-14;  
 Matches 38; Conservative 11; Mismatches 28; Indels 4; Gaps 2;  
 OY 1 MRLVLSLLCILLCFSTFTEGRRPAKMSGRRTLCCHRVSPNSTNLKSHVRLC 60  
 Db 1 MRLTTLGLFMFLCLCVLSSEGRKRPAPK---PKLRPCHLSRSPKPTWKGNTKRPC 57  
 OY 61 KPCLEPEPRRLMYYPGALPOV 81  
 Db 58 RPCR-KUESNSWVVPALPOI 77  
 RESULT 4  
 AAE16485  
 ID AAE16485 standard; peptide; 30 AA.  
 AC AAE16485;  
 DT 09-APR-2002 (first entry)  
 DE Human secreted epithelial colon stromal-1 (Secs-1) peptide #2.  
 XX Secreted epithelial colon stromal-1; Secs-1; gene therapy; osteopetrosis;  
 KW hematopoietic disorder; osteoporosis; osteogenesis imperfecta; cachexia;  
 KW Paget's disease; periodontal disease; hypercalcaemia; glomerulonephritis;  
 KW diabetes; obesity; osteopetritic; cystostatic; nephrotrophic; antidiabetic;  
 KW anorectic; immunomodulator; antipsoriatic; vulnery; antinfertility;  
 KW gynaecological; antitumor; antinflammatory; cancer; cell therapy;  
 KW human.  
 XX Homo sapiens.  
 OS  
 XX  
 PN WO200198497-A1.  
 PD 27-DEC-2001.  
 XX 28-NOV-2000; 2000WO-US32479.  
 PF 21-JUN-2000; 2000US-0599087.  
 PR 28-NOV-2000; 2000US-0724000.  
 XX (AMGE-) AMGEN INC.  
 PA Polverino AJ, Luethy R;  
 PI  
 XX Secreted epithelial colon stromal-1 polypeptides and nucleic acids,

DR WPI: 2002-122281/16.  
 DR N-PSDB; AAD27026.  
 XX Secreted epithelial colon stromal-1 polypeptides and nucleic acids,  
 PT useful for diagnosing, treating and preventing hematopoietic disorder,  
 PT osteoporosis, Paget's disease, cancer, diabetes -  
 XX Disclosure; Fig 4D; 134pp; English.  
 PS  
 XX The present invention relates to an isolated murine or human secreted  
 CC epithelial colon stromal-1 (Secs-1) polypeptide, its allelic or splice  
 CC variant, orthologue, fragment or mutant. Secs-1 gene is used in gene  
 CC therapy and cell therapy. Secs-1 is useful for identifying a compound  
 CC which binds to a Secs-1 polypeptide. Secs-1 is useful for treating,  
 CC preventing or ameliorating a disease condition such as hematopoietic  
 CC disorder, osteoporosis, osteopetrosis, osteogenesis imperfecta, Paget's  
 CC disease, periodontal disease, hypercalcaemia, acute glomerulonephritis,  
 CC chronic glomerulonephritis, cancer, diabetes, obesity or cachexia. Secs-1  
 CC is also useful for diagnosing a pathological condition which involves  
 CC determining the presence or amount of Secs-1 or polypeptide encoded by  
 CC Secs-1 DNA in a sample; and diagnosing a pathological condition, or  
 CC susceptibility to pathological condition based on the presence or amount  
 CC of expression of the polypeptide. The present sequence is human Secs-1  
 CC peptide encoded by second exon.  
 CC  
 XX Sequence 30 AA;  
 SQ  
 Query Match 38.2%; Score 169; DB 23; Length 30;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-12;  
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 24 GKRRPAKMSGRRTLCCHRVSPNSTNLK 53  
 Db 1 GKRRPAKMSGRRTLCCHRVSPNSTNLK 30  
 RESULT 5  
 AAE16486  
 ID AAE16486 standard; peptide; 28 AA.  
 AC AAE16486;  
 DT 09-APR-2002 (first entry)  
 DE Human secreted epithelial colon stromal-1 (Secs-1) peptide #3.  
 XX Secreted epithelial colon stromal-1; Secs-1; gene therapy; osteopetrosis;  
 KW hematopoietic disorder; osteoporosis; osteogenesis imperfecta; cachexia;  
 KW Paget's disease; periodontal disease; hypercalcaemia; glomerulonephritis;  
 KW diabetes; obesity; osteopetritic; cystostatic; nephrotrophic; antidiabetic;  
 KW anorectic; immunomodulator; antipsoriatic; vulnery; antinfertility;  
 KW gynaecological; antitumor; antinflammatory; cancer; cell therapy;  
 KW human.  
 XX Homo sapiens.  
 OS  
 XX  
 PN WO200198497-A1.  
 PD 27-DEC-2001.  
 XX 28-NOV-2000; 2000WO-US32479.  
 PF 21-JUN-2000; 2000US-0599087.  
 PR 28-NOV-2000; 2000US-0724000.  
 XX (AMGE-) AMGEN INC.  
 PA Polverino AJ, Luethy R;  
 PI  
 XX WPI: 2002-122281/16.  
 DR N-PSDB; AAD27026.  
 PT Secreted epithelial colon stromal-1 polypeptides and nucleic acids,

PT useful for diagnosing, treating and preventing hematopoietic disorder,  
 PT osteoporosis, Paget's disease, cancer, diabetes -  
 XX  
 PS Disclosure; Fig 4F; 134pp; English.  
 CC The present invention relates to an isolated murine or human secreted  
 CC epithelial colon stromal-1 (Secs-1) polypeptide, its allelic or splice  
 CC variant, orthologue, fragment or mutant. Secs-1 gene is used in gene  
 CC therapy and cell therapy. Secs-1 is useful for identifying a compound  
 CC which binds to a Secs-1 polypeptide. Secs-1 is useful for treating,  
 CC preventing or ameliorating a disease condition such as hematopoietic  
 CC disorder, osteoporosis, osteopetrosis, osteogenesis imperfecta, Paget's  
 CC disease, periodontal disease, hypercalcaemia, acute glomerulonephritis,  
 CC chronic glomerulonephritis, cancer, diabetes, obesity or cachexia. Secs-1  
 CC is also useful for diagnosing a pathological condition which involves  
 CC determining the presence or amount of Secs-1 or polypeptide encoded by  
 CC Secs-1 DNA in a sample; and diagnosing a pathological condition, or  
 CC susceptibility to pathological condition based on the presence or amount  
 CC of expression of the polypeptide. The present sequence is human Secs-1  
 CC peptide encoded by third exon.  
 CC  
 XX  
 SQ Sequence 28 AA;  
 Query Match 36.9%; Score 163; DB 23; Length 28;  
 Best Local Similarity 100.0%; Pred. No. 5, 1e-12;  
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 54 GHVRLCKPCKLEPEPRMLVVGALPOV 81  
 ID 1 GHVRLCKPCKLEPEPRMLVVGALPOV 28  
 Db 1 GHVRLCKPCKLEPEPRMLVVGALPOV 28  
 RESULT 6  
 ID AAE16480 standard; peptide; 54 AA.  
 XX  
 AC AAE16480;  
 XX  
 DT 09-APR-2002 (first entry)  
 XX  
 DE Mouse Secreted epithelial colon stromal-1 (Secs-1) protein fragment.  
 XX  
 KW Secreted epithelial colon stromal-1; Secs-1; gene therapy; osteopetrosis;  
 KW hematopoietic disorder; osteoporosis; osteogenesis imperfecta; cachexia;  
 KW Paget's disease; periodontal disease; hypercalcaemia; glomerulonephritis;  
 KW diabetes; obesity; osteopathic; cytostatic; nephrotrophic; antidiabetic;  
 KW anorectic; immunomodulator; antipsoriatic; vulnery; antiinfectility;  
 KW gynaecological; antiulcer; antiinflammatory; cancer; cell therapy;  
 KW mouse.  
 KM  
 XX  
 OS Mus musculus.  
 XX  
 PN WO200198497-A1.  
 XX  
 PD 27-DEC-2001.  
 XX  
 PF 28-NOV-2000; 2000MO-US32479.  
 XX  
 PR 21-JUN-2000; 2000US-0599087.  
 XX  
 PR 28-NOV-2000; 2000US-0724000.  
 XX  
 PA (AMGE-) AMGEN INC.  
 XX  
 PI Polverino AJ, Luethy R;  
 XX  
 DR WPI; 2002-122281/16.  
 XX  
 PT Secreted epithelial colon stromal-1 polypeptides and nucleic acids,  
 PT useful for diagnosing, treating and preventing hematopoietic disorder,  
 PT osteoporosis, Paget's disease, cancer, diabetes -  
 XX  
 PS Claim 14; Page 120; 134pp; English.

CC The present invention relates to an isolated murine or human secreted  
 CC epithelial colon stromal-1 (Secs-1) polypeptide, its allelic or splice  
 CC variant, orthologue, fragment or mutant. Secs-1 gene is used in gene  
 CC therapy and cell therapy. Secs-1 is useful for identifying a compound  
 CC which binds to a Secs-1 polypeptide. Secs-1 is useful for treating,  
 CC preventing or ameliorating a disease condition such as hematopoietic  
 CC disorder, osteoporosis, osteopetrosis, osteogenesis imperfecta, Paget's  
 CC disease, periodontal disease, hypercalcaemia, acute glomerulonephritis,  
 CC chronic glomerulonephritis, cancer, diabetes, obesity or cachexia. Secs-1  
 CC is also useful for diagnosing a pathological condition which involves  
 CC determining the presence or amount of Secs-1 or polypeptide encoded by  
 CC Secs-1 DNA in a sample; and diagnosing a pathological condition, or  
 CC susceptibility to pathological condition based on the presence or amount  
 CC of expression of the polypeptide. The present sequence is mouse Secs-1  
 CC protein fragment.  
 CC  
 XX  
 SQ Sequence 54 AA;  
 Query Match 30.0%; Score 132.5; DB 23; Length 54;  
 Best Local Similarity 45.6%; Pred. No. 3, 9e-08;  
 Matches 26; Conservative 8; Mismatches 20; Indels 3; Gaps 1;  
 QY 25 KRPAKAMSGRRTRLCCHRVSPNSTNLKGHVRLCKPCKLEPEPRMLVVGALPOV 81  
 ID 1 KRPAKAMSGRRTRLCCHRVSPNSTNLKGHVRLCKPCKLEPEPRMLVVGALPOV 54  
 Db 1 KRPAKAMSGRRTRLCCHRVSPNSTNLKGHVRLCKPCKLEPEPRMLVVGALPOV 54  
 RESULT 7  
 ID AAU45851 standard; Protein; 78 AA.  
 XX  
 AC AAU45851;  
 XX  
 DT 27-FEB-2002 (first entry)  
 XX  
 DE Propionibacterium acnes immunogenic protein #6747.  
 XX  
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 KM  
 XX  
 OS Propionibacterium acnes.  
 XX  
 PN WO200181581-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 20-APR-2001; 2001MO-US12865.  
 XX  
 PR 21-APR-2000; 2000US-199047P.  
 XX  
 PR 02-JUN-2000; 2000US-208841P.  
 XX  
 PR 07-JUL-2000; 2000US-216747P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Skelly YAW, Persing DH, Mitcham JU, Wang SS, Bhatia A;  
 XX  
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 XX  
 DR WPI; 2001-616774/71.  
 XX  
 DR N-PSDB; AAS59528.  
 XX  
 PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris -  
 XX  
 PS Example 1; SEQ ID No 7046; 1069pp; English.  
 XX  
 CC Sequences AAU39105-AA68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.

P. acnes is also involved in infections of bone, joints and the central nervous system, however it is particularly involved in the inflammatory lesions associated with acne vulgaris. A method for detecting the presence or absence of P. acnes in a patient comprises contacting a sample with a binding agent that binds to the proteins of the invention and determining the amount of bound protein in the sample. The polypeptides may be used as antigens in the production of antibodies specific for P. acnes proteins. These antibodies can be used to downregulate expression and activity of P. acnes polypeptides and therefore treat P. acnes infections. The antibodies may also be used as diagnostic agents for determining P. acnes presence, for example, by enzyme linked immunosorbent assay (ELISA).

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pcr\_sequences.

SO Sequence 78 AA;

Query Match 14.0%; Score 62; DB 22; Length 78;  
Best Local Similarity 78.6%; Pred. No. 9.9; Mismatches 1; Conservative 1; Indels 0; Gaps 0;

Db 25 KRRPAAKSGRRR 38  
50 RRRPRAKSGRRR 63

RESULT 8

AAU66859  
ID AAU66859 standard; Protein; 75 AA.

AC AAU66859;

DT 13-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #27755.

KM SAPHO syndrome; synovitis; acne; pustulosis; hypertostis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

PN WO200181581-A2.

PD 01-NOV-2001.

PF 20-APR-2001; 2001MO-US12865.

PR 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

XX Skelky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI I'maisonneuve J, Zhang Y, Jen S, Carter D;

DR WPI: 2001-616774/71.  
DR N-PSDB; AAS59761.

XX Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris -

PS Example 1; SEQ ID No 28054; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertostis and osteomyelitis), uveitis and endophthalmitis.

P. acnes is also involved in infections of bone, joints and the central nervous system, however it is particularly involved in the inflammatory lesions associated with acne vulgaris. A method for detecting the presence or absence of P. acnes in a patient comprises contacting a sample with a binding agent that binds to the proteins of the invention and determining the amount of bound protein in the sample. The polypeptides may be used as antigens in the production of antibodies specific for P. acnes proteins. These antibodies can be used to downregulate expression and activity of P. acnes polypeptides and therefore treat P. acnes infections. The antibodies may also be used as diagnostic agents for determining P. acnes presence, for example, by enzyme linked immunosorbent assay (ELISA).

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pcr\_sequences.

SO Sequence 75 AA;

Query Match 13.8%; Score 61; DB 22; Length 75;  
Best Local Similarity 34.3%; Pred. No. 12; Mismatches 12; Conservative 5; Indels 18; Gaps 0;

Db 41 CHRVPSPSTNLGHHVRLCKPCKLEPEPLWVP 75  
32 CRSVASSSTASAKGWSLALQPCPLRPLTPVSWCRP 66

RESULT 9

AAU54516  
ID AAU54516 standard; Protein; 75 AA.

AC AAU54516;

DT 27-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #15412.

KM SAPHO syndrome; synovitis; acne; pustulosis; hypertostis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

PN WO200181581-A2.

PD 01-NOV-2001.

PF 20-APR-2001; 2001MO-US12865.

PR 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

XX Skelky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI I'maisonneuve J, Zhang Y, Jen S, Carter D;

DR WPI: 2001-616774/71.  
DR N-PSDB; AAS59565.

XX Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris -

PS Example 1; SEQ ID No 15711; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertostis and osteomyelitis), uveitis and endophthalmitis.

P. acnes is also involved in infections of bone, joints and the central nervous system, however it is particularly involved in the inflammatory lesions associated with acne vulgaris. A method for detecting the presence or absence of P. acnes in a patient comprises contacting a sample with a binding agent that binds to the proteins of the invention and determining the amount of bound protein in the sample. The polypeptides may be used as antigens in the production of antibodies specific for P. acnes proteins. These antibodies can be used to downregulate expression and activity of P. acnes polypeptides and therefore treat P. acnes infections. The antibodies may also be used as diagnostic agents for determining P. acnes presence, for example, by enzyme linked immunosorbent assay (ELISA).

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [http://www.int/pub/published\\_pct\\_sequences](http://www.int/pub/published_pct_sequences).

Sequence 75 AA:

Query Match 13.8%; Score 61; DB 22; Length 75;  
Best Local Similarity 34.3%; Pred. No. 12;  
Matches 12; Conservative 5; Mismatches 18; Indels 0; Gaps 0;

QY 41 CHRVSPNSTLKGHVRLCKPCKLEPEPLWVP 75  
DB 32 CRVASSSTASAKGMSIALCOPCLRPPLTPVSMCRP 66

RESULT 10  
AAG00379  
ID AAG00379 standard; Protein: 73 AA.

AC AAG00379;

DT 06-OCT-2000 (first entry)

DE Human secreted protein, SEQ ID NO: 4460.

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation; gene therapy; chromosome mapping.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

PA (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI: 2000-500381/45.

DR N-PSDB: AAC00385.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for diagnostic, forensic, gene therapy and chromosome mapping procedures -

PS Claim 13; SEQ ID 4460; 71pp + CD-ROM; English.

The present sequence is a polypeptide encoded by one of a large number of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs derived from 30 different tissues. EST sequences usually correspond mainly to the 3' untranslated region (UTR) of the mRNA because they are often obtained from oligo-dT primed cDNA libraries. Such ESTs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs and even in those cases where longer cDNA sequences have been obtained, the full 5' UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be used to obtain full length cDNAs and genomic

CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and CC chromosome mapping procedures. They are used to obtain upstream CC regulatory sequences and to design expression and secretion vectors.

Sequence 73 AA:

Query Match 13.5%; Score 59.5; DB 21; Length 73;  
Best Local Similarity 26.8%; Pred. No. 18;  
Matches 22; Conservative 7; Mismatches 16; Indels 37; Gaps 4;

QY 2 RLVLVLS-----SLCLILCFSTFTEGKRRRAKMSGRRLCCHRVSPNSTLNG 54  
DB 12 QLVLMHCVLFCFVLFCEVLFCEVLFET-GSQSYAQA----- 47

QY 55 HHVRLCKPCKLEPEPLWVP 76  
DB 48 -GVQSCDCLQPPP-----PG 63

RESULT 11  
AAV36266  
ID AAV36266 standard; Protein: 77 AA.

AC AAV36266;

DT 17-SEP-1999 (first entry)

DE Human secreted protein encoded by gene 43.

KW Human; secreted protein; cancer; tumour; developmental abnormality; foetal deficiency; blood disorder; immune system disorder; inflammation; autoimmune disease; allergy; Alzheimer's disease; cognitive disorder; schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder; atherosclerosis; diabetes; cardiovascular disorder; kidney disorder; digestive disorder; endocrine disorder; infection; AIDS.

OS Homo sapiens.

PN WO9931117-A1.

PD 24-JUN-1999.

PF 17-DEC-1998; 98WO-US27059.

PR 19-DEC-1997; 97US-0068369.

PR 18-DEC-1997; 97US-0068006.

PR 18-DEC-1997; 97US-0068007.

PR 18-DEC-1997; 97US-0068008.

PR 18-DEC-1997; 97US-0068053.

PR 18-DEC-1997; 97US-0068054.

PR 18-DEC-1997; 97US-0068057.

PR 18-DEC-1997; 97US-0068064.

PR 18-DEC-1997; 97US-0070923.

PR 19-DEC-1997; 97US-0068169.

PR 19-DEC-1997; 97US-0068365.

PR 19-DEC-1997; 97US-0068367.

PR 19-DEC-1997; 97US-0068368.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Carter KC, Duan RD, Feng P, Ferrie AM, Florence C;

PI Florence K, Greene JM, Janat F, Kyaw H, Moore PA;

PI Ni J, Rosen CA, Ruben SM, Shi Y, Soppet DR, Wei Y;

PI Yu G;

DR WPI: 1999-418749/35.

DR N-PSDB: AAX97958.

PT New isolated human genes encoding secreted polypeptides

PS Claim 11; Page 368; 537pp; English.

AXX97916 to AAX98029 represent 110 isolated human secreted protein



PR 22-AUG-1997; 9705-0056911.  
 PR 05-SEP-1997; 9705-0057650.  
 PR 05-SEP-1997; 9705-0057761.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Bednarek DP, Brewer LA, Carter KC, Duan R, Edner R, Endress GA;  
 PI Feng P, Ferlie AM, Fischer CL, Graves KA, Greene JM, Hu JS;  
 PI Kyaw H, Lafleur DM, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;  
 PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;  
 XX  
 DR WPI: 1998-609887/51.  
 DR N-PSDB; AAV34275.  
 XX  
 PT New isolated human genes and the secreted polypeptides they encode  
 PT - useful for diagnosis and treatment of e.g. cancers, neurological  
 PT disorders, immune diseases, inflammation or blood disorders  
 XX  
 PS Claim 1; Page 352; 447pp; English.  
 XX  
 CC This sequence represents a secreted human protein encoded by the gene  
 CC clone detailed in the descriptor line.  
 CC The gene can be used to generate fusion proteins by linking to the gene  
 CC to a human immunoglobulin Fc portion (e.g. AAV34145) for increasing the  
 CC stability of the fused protein as compared to the human protein only.  
 CC The invention relates to 70 novel genes and their fragments (nucleic  
 CC acid sequences: AAV34154-V34276; amino acid sequences AAV75057-W75179)  
 CC which are useful for preventing, treating or ameliorating medical  
 CC conditions e.g. by protein or gene therapy. Also, pathological  
 CC conditions can be diagnosed by determining the amount of the new  
 CC polypeptides in a sample or by determining the presence of mutations in  
 CC the new polynucleotides. Specific uses are described for each of the 70  
 CC polynucleotides, based on which tissues they are most highly expressed in  
 CC (see AAV34154 for described uses).  
 XX  
 SQ Sequence 53 AA;  
 Query Match 13.1%; Score 58; DB 19; Length 53;  
 Best Local Similarity 50.0%; Pred. No. 19;  
 Matches 12; Conservative 6; Mismatches 6; Indels 0; Gaps 0;  
 QY 3 LTVSSLLCILLCFSTFSTGKR 26  
 DB 19 LLLIASLAPLWLYCWRMPTKGR 42  
 XX  
 RESULT 13  
 ID AAO05651 standard; Protein: 62 AA.  
 AC AAO05651;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human polypeptide SEQ ID NO 19543.  
 XX  
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; Inflammation.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200164835-A2.  
 XX  
 PD 07-SEP-2001.  
 XX  
 PF 26-FEB-2001; 2001WO-US04927.  
 XX  
 PR 28-FEB-2000; 2000US-0515126.  
 PR 18-MAY-2000; 2000US-0577409.  
 XX  
 PA (HYSE-) HYSEQ INC.

XX  
 PI Tang YT, Liu C, Drmanac RT;  
 XX  
 DR WPI: 2001-514838/56.  
 DR N-PSDB; AAI85582.  
 XX  
 PT Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders -  
 XX  
 PS Claim 20; SEQ ID NO 19543; 1399pp + Sequence Listing; English.  
 XX  
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activity/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pot\_sequences.  
 XX  
 SQ Sequence 62 AA;  
 Query Match 13.1%; Score 58; DB 22; Length 62;  
 Best Local Similarity 28.6%; Pred. No. 23;  
 Matches 18; Conservative 5; Mismatches 30; Indels 10; Gaps 3;  
 QY 10 LCILLCFSTFSTGKR--RPKAMSGRRRLCCHRVSPNSTNKK-GHYRLCKPKLE 66  
 DB 2 LCILLCFCCYFETRSRVAGVGMHD-----HSPQPPPRPKSSHLSTRHDS 54  
 XX  
 QY 67 PEP 69  
 DB 55 PEP 57  
 XX  
 RESULT 14  
 ID AAB44937 standard; Protein: 63 AA.  
 AC AAB44937;  
 XX  
 DT 12-FEB-2001 (first entry)  
 XX  
 DE Human secreted protein encoded by gene 21.  
 XX  
 KW Secreted protein; human; immunosuppressive; antiarthritic; antirheumatic;  
 KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;  
 KW nootropic; neuroprotective; antibacterial; vitinide; fungicide;  
 KW ophthalmological; gene therapy; treatment; autoimmune disease; infection;  
 KW cardiovascular disorder; nervous system disorder; ocular disorder;  
 KW wound healing; epithelial cell proliferation; skin aging; mental state;  
 KW transplantation; metabolism modulation.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200055200-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 09-MAR-2000; 2000NO-US06042.  
 XX  
 PR 12-MAR-1999; 99US-0124143.  
 PR 03-DEC-1999; 99US-0168663.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM, Komatsoulis G;  
 XX WPI; 2000-656008/63.  
 DR N-PSDB; AAC79918.  
 XX  
 XX Isolated human secretory proteins, nucleic acids encoding them and  
 PT antibodies directed against them, useful for diagnosing and treating  
 PT disorders related to the proteins such as cancer, Alzheimer's disease  
 PT and Parkinsons -  
 XX  
 PS Claim 11; Page 388; 453pp; English.  
 XX  
 XX This invention describes a novel isolated polypeptide (I) and its  
 CC encoding nucleic acid molecule (II) which have immunosuppressive,  
 CC antiarthritic, antirheumatic, antiproliferative, cytostatic, cardiant,  
 CC vasotropic, cerebroprotective, neurotropic, neuroprotective, antibacterial,  
 CC virocidic, fungicide and ophthalmological activity and which can be used  
 CC for gene therapy. (I) and (II) are used to prevent, treat or ameliorate  
 CC a medical condition in e.g. humans, mice, rabbits, goats, horses, cats,  
 CC dogs, chickens or sheep. (I) and (II) are also used in diagnosing a  
 CC pathological condition or susceptibility to a pathological condition. The  
 CC antibodies to (II) can also be used in alleviating symptoms associated  
 CC with the disorders and in diagnostic immunoassays e.g. radioimmunoassays  
 CC or enzyme linked immunoassay assays (ELISA). Disorders which are  
 CC diagnosed or treated include autoimmune diseases e.g. Rheumatoid  
 CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or  
 CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular  
 CC disorders e.g. cerebral ischemia, angogenesis, nervous system disorders  
 CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and  
 CC fungi and ocular disorders e.g. corneal infection. The polypeptides can  
 CC also be used to aid wound healing and epithelial cell proliferation, to  
 CC prevent skin aging due to sunburn, to maintain organs before  
 CC transplantation, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The polypeptides are used to  
 CC modulate mammalian metabolism, to change mammal's mental state or  
 CC physical state by influencing biorhythms circadian rhythms, depression  
 CC tendency for violence tolerance for pain, reproductive capabilities,  
 CC hormonal or endocrine levels, appetite, libido, memory, stress or other  
 CC cognitive qualities, as a food additive or preservative, such as to  
 CC increase or decrease storage capabilities, fat content, lipid, protein,  
 CC carbohydrates, vitamins, minerals, cofactors or other nutritional  
 CC components.  
 XX  
 XX Sequence 63 AA:  
 SQ  
 Query Match 13.1%; Score 58; DB 21; Length 63;  
 Best Local Similarity 30.4%; Pred. No. 23;  
 Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;  
 OY 2 RLIVSSLCITLLCFSTFGKRRPAKAWSGRRTRLC-CHRVSPNSTNLKGHHVRLC 60  
 DB 5 RLIGTSSFFCITL-----GLAWTAPASESCGPHPLAEPSTVILG-AIFPC 49  
 OY 61 KPCKLEPP 69  
 DB 50 RTGSLSPAP 58  
 RESULT 15  
 AAB60725  
 ID AAB60725 standard: protein; 63 AA.  
 XX  
 AC AAB60725;  
 XX  
 DT 27-MAR-2001 (first entry)  
 XX  
 DE Human secreted protein #24.  
 XX  
 XX Secreted protein; gene therapy; vaccine; cancer; leukemia;  
 KW autoimmune disease; allergy; inflammation; graft rejection;  
 KW hyperproliferation; cardiovascular; infection.  
 XX  
 OS Homo sapiens.

XX  
 PN WO200076531-A1.  
 PD 21-DEC-2000.  
 XX  
 XX 01-JUN-2000; 2000WO-US15137.  
 PF 11-JUN-1999; 99US-0138625.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX  
 PI Rosen CA, Ruben SM, Komatsoulis GA;  
 XX WPI; 2001-071148/08.  
 DR  
 XX  
 PT Nucleic acids encoding 47 human secreted polypeptides, useful for  
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's  
 PT disease and diabetic retinopathy -  
 XX  
 PS Claim 11; Page 485; 525pp; English.  
 XX  
 XX The present invention relates to 26 secreted human proteins. The  
 CC proteins may be used in the prevention, diagnosis and treatment of  
 CC diseases associated with inappropriate polypeptide expression.  
 CC For example, they may be used in gene therapy or in vaccines.  
 CC Typical of diseases which are potentially treatable are cancers  
 CC (including leukemia), autoimmune diseases, allergies, inflammation,  
 CC graft rejection, hyperproliferation, cardiovascular diseases  
 CC (particularly critical limb ischemia and coronary disease) and any  
 CC involving abnormal angiogenesis, neurodegeneration and/or  
 CC infectious diseases.  
 CC  
 XX Sequence 63 AA:  
 SQ  
 Query Match 13.1%; Score 58; DB 22; Length 63;  
 Best Local Similarity 30.4%; Pred. No. 23;  
 Matches 21; Conservative 8; Mismatches 24; Indels 16; Gaps 3;  
 OY 2 RLIVSSLCITLLCFSTFGKRRPAKAWSGRRTRLC-CHRVSPNSTNLKGHHVRLC 60  
 DB 5 RLIGTSSFFCITL-----GLAWTAPASESCGPHPLAEPSTVILG-AIFPC 49  
 OY 61 KPCKLEPP 69  
 DB 50 RTGSLSPAP 58  
 Search completed: May 8, 2003, 21:46:30  
 Job time : 36 secs



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 8, 2003, 21:47:43 ; Search time 1229 Seconds  
(Without alignments)  
1918.086 Million cell updates/sec

Title: US-09-599-087B-5  
Perfect score: 442  
Sequence: 1 MRLVLSLLCILLICRSIF.....PKRLEPPRLMVPGALPQV 81

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 463318

Minimum DB seq length: 25  
Maximum DB seq length: 80

Post-processing: Minimum Match 0%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp  
-O=/cgn2.1/USPRO.spool/US09599087/runat.07052003.113321.7344/app.query.fasta\_1.263  
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCLIN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=25 -MAXLEN=80  
-USER=US09599087.0CGN.1.1.1319.0runat.07052003.113321.7344 -NCPU=6 -ICPU=3  
-NO.XLPXY -NO.MMAP -LARGOQUERY -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV.TIMEOUT=10 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pt:\*  
10: gb\_ro:\*  
11: gb\_sy:\*  
12: gb\_un:\*  
13: gb\_vl:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sy:\*  
28: em\_un:\*

29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rdi:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	59.5	13.5	76	6	A07238
2	56	12.7	42	6	AX342232
3	53	12.0	79	6	I86930
4	51.5	11.7	69	9	HSU90998
5	49	11.1	66	10	MUSCAY
6	49	11.1	70	10	RNTCRAJ9
7	48	10.9	76	9	S75385S18
8	47	10.6	51	6	AX161822
9	46.5	10.5	48	6	AR077574
10	46	10.4	71	6	AX025072
11	46	10.4	71	6	AX025094
12	46	10.4	72	6	I71481
13	46	10.4	78	9	HSIGKAB01
14	45.5	10.3	69	6	AR005307
15	45	10.2	39	6	AR077518
16	45	10.2	48	6	AX160227
17	45	10.2	76	9	AF118405
18	44	10.0	42	6	AR014091
19	44	10.0	50	6	I28909
20	44	10.0	50	6	I28910
21	44	10.0	51	6	AR077572
22	44	10.0	51	6	AR077575
23	44	10.0	66	6	A28858
24	44	10.0	79	5	DUKAEFC04
25	43	9.7	46	6	A06605
26	43	9.7	46	6	A10221
27	43	9.7	51	6	AX159330
28	43	9.7	66	6	AX040385
29	43	9.7	72	9	HSADD2S31
30	43	9.7	76	1	MCYTRMF
31	43	9.7	76	1	MCYTRMF
32	43	9.7	76	6	A07239
33	43	9.7	78	6	AR000573
34	43	9.7	78	6	AR000574
35	43	9.7	78	6	AR208886
36	43	9.7	78	6	AR208887
37	43	9.7	78	6	BD003625
38	43	9.7	78	6	BD003626
39	43	9.7	79	6	AR050027
40	42.5	9.6	69	9	HSU91209
41	42	9.5	30	6	AR022280
42	42	9.5	47	6	AX003724
43	42	9.5	51	6	AX159329
44	42	9.5	51	6	AX165793
45	42	9.5	52	9	HS19

## ALIGNMENTS

RESULT 1

```

A07238
LOCUS       A07238               76 bp    DNA             linear   PAT 24-AUG-1993
DEFINITION   Probe for vpa-beta.
ACCESSION   A07238
VERSION     A07238.1  GI:413006
KEYWORDS    SOURCE              synthetic construct.
            ORGANISM            synthetic construct.
            REFERENCE            artificial sequences.
            AUTHORS              1 (bases 1 to 76)
                                   Baldus, B., Donner, P., Schleuning, W.D., Alagon, A., Boldol, W.,
                                   Kraetzschmar, J.R., Haendler, B.J. and Langer, G.
            TITLE                Novel thrombolytic
            JOURNAL              Patent: EP 0383417-A 1 22-AUG-1990;
            FEATURES              Schering Aktiengesellschaft
            source                Location/Qualifiers
            BASE COUNT           14 a 17 c 24 g 21 t
            ORIGIN
Alignment Scores:
Pred. No.:      566              Length:      76
Score:          59.50            Matches:     14
Percent Similarity: 66.67%      Conservative: 4
Best Local Similarity: 51.85%    Mismatches: 7
Query Match:    13.46%          Indels:     2
DB:             6               Gaps:       1

US-09-599-087B-5 (1-81) x A07238 (1-76)
QY          9  LeuLeuGysAlleLeuLeuLeuGysPheSerIlePheSer-ThrGluGlyLysArgArgPr 28
Db          1  CTGTTGTGTACTGCTGCTGTGGAGCAGCTCTCTCGTTGCCAGGCC---AGGAACCC 57

QY          28  cAlaLysAlaTrpSerGly 34
Db          58  TACAGGCATTTGCCAGGG 76

RESULT 2
AX342232
LOCUS       AX342232               42 bp    DNA             linear   PAT 12-JAN-2002
DEFINITION   Sequence 18 from Patent WO0198497.
ACCESSION   AX342232
VERSION     AX342232.1  GI:18151770
KEYWORDS    SOURCE              synthetic construct.
            ORGANISM            synthetic construct.
            REFERENCE            artificial sequences.
            AUTHORS              1
            TITLE                Polyvino, A.J. and Luethy, R.
            JOURNAL              Secreted epithelial colon stromal-1 polypeptides, nucleic acids
            FEATURES              encoding the same and uses thereof
            source                Patent: WO 0198497-A 18 27-DEC-2001;
                                   Amgen, Inc. (US)
                                   Location/Qualifiers
                                   1..42
                                   /organism="synthetic construct"
                                   /db_xref="taxon:32630"
                                   /note="PCR primer corresponding to human SECS-1"
BASE COUNT   15 a 10 c 9 g 8 t
ORIGIN
Alignment Scores:
Pred. No.:      642              Length:      42
Score:          56.00            Matches:     10
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%    Mismatches: 0
Query Match:    12.67%          Indels:     0
DB:             6               Gaps:       0

US-09-599-087B-5 (1-81) x AX342232 (1-42)

```

---

```

QY          25  LysArgArgProAlaLysAlaTrpSerGly 34
Db          13  AAACGTCTCCAGCTAAAGCTGTCTCAGGC 42

RESULT 3
LOCUS       186930               79 bp    DNA             linear   PAT 10-JUN-1998
DEFINITION   Sequence 19 from patent US 5702931.
ACCESSION   186930
VERSION     186930.1  GI:3206648
KEYWORDS    SOURCE              Unknown.
            ORGANISM            Unclassified.
            REFERENCE            1 (bases 1 to 79)
            AUTHORS              Andrews, W.R., Morser, M.J. and Vilander, L.R.
            TITLE                Mutagenesis methods and compositions
            JOURNAL              Patent: US 5702931-A 19 30-DEC-1997;
            FEATURES              Location/Qualifiers
            source                1..79
                                   /organism="unknown"
BASE COUNT   10 a 28 c 31 g 10 t
ORIGIN
Alignment Scores:
Pred. No.:      2,45e+03          Length:      79
Score:          53.00            Matches:     9
Percent Similarity: 55.00%      Conservative: 2
Best Local Similarity: 45.00%    Mismatches: 9
Query Match:    11.99%          Indels:     0
DB:             6               Gaps:       0

US-09-599-087B-5 (1-81) x 186930 (1-79)
QY          29  AlAlaLysAlaTrpSerGlyArgArgThrArgLeuGysCysHisArgValProSerProAsn 48
Db          15  GCCTCGCGCTGGCGGGGGCTCCGACAGCGCGTGTCCACCGTCCACCTGTGCCAGT 74

RESULT 4
HSU90998/c
LOCUS       HSU90998               69 bp    mRNA             linear   PRI 15-JUL-1997
DEFINITION   Homo sapiens clone FEL389 T-cell receptor delta chain (TCRDVLJ2)
ACCESSION   U90998
VERSION     U90998.1  GI:2239403
KEYWORDS    SOURCE              Homo sapiens.
            ORGANISM            Homo sapiens.
            REFERENCE            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            AUTHORS              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
            TITLE                1 (bases 1 to 69)
            JOURNAL              Holtmeier, W., Witthoft, T., Hennemann, A., Winter, H.S. and
            FEATURES              Kagnoff, M.F.
                                   The TCR-delta repertoire in human intestine undergoes
                                   characteristic changes during fetal to adult development
                                   J. Immunol. 158 (12), 5632-5641 (1997)
            MEDLINE              97334214
            PUBMED              9190911
            REFERENCE            2 (bases 1 to 69)
            AUTHORS              Holtmeier, W., Witthoft, T., Hennemann, A., Harland, S.W. and
            TITLE                Kagnoff, M.F.
            JOURNAL              Direct Submussion
            FEATURES              Submitted (27-FEB-1997) Department of Medicine, University of
                                   Frankfurt, Theodor-Stern Kai #7, Frankfurt 60590, Germany
                                   Location/Qualifiers
                                   1..69
                                   /organism="Homo sapiens"
                                   /db_xref="taxon:9606"
                                   /chromosome="14"
                                   /clone="FEL389"
                                   /tissue-type="small intestine"
                                   /dev-stage="fetus; 20 weeks gestation"

```

gene  
 <1..>69  
 /gene="TCRDVIJ2"  
 CDS  
 <1..>69  
 /gene="TCRDVIJ2"  
 /note="rearranged: contains CDR3 domain, 11 amino acids (calculation according to: Rock, E., J. Exp. Med., 179:323-328, 1994)"  
 /product="T-cell receptor delta chain"  
 /codon\_start=1  
 /protein\_id="AAC51561.1"  
 /db\_xref="GI:239404"  
 /translation="IEDSAKFKCALGELGIVTAOLF"  
 V\_region  
 <1..>69  
 /gene="TCRDVIJ2"  
 BASE COUNT 18 a 14 c 19 g 18 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 2.95e+03 Length: 69  
 Score: 51.50 Matches: 9  
 Percent Similarity: 73.33% Conservative: 2  
 Best Local Similarity: 60.00% Mismatches: 3  
 Query Match: 11.65% Indels: 1  
 DB: 9 Gaps: 1

US-09-599-087b-5 (1-81) x HSU90998 (1-69)

OY 39 LeucyCyvHlaRgVAlProSeR---ProAnSeRThraNleu. 52  
 |||||  
 Db 63 TTGTGCTGTACCGTATCCCAAGTTCACCAAGACACAAAGTA 19

RESULT 5  
 MUSTCAYJ 66 bp mRNA linear ROD 27-APR-1993  
 LOCUS  
 DEFINITION Mouse T-cell receptor alpha-chain mRNA J-C region, partial cds.  
 ACCESSION M38677.1 GI:201214  
 VERSION M38677.1 J-region; T-cell receptor.  
 KEYWORDS C-region; J-region; T-cell receptor.  
 SOURCE Mouse, CDNA to mRNA.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.  
 Yague, J., Blackman, M., Born, W., Marreck, P., Kappler, J. and Palmer, E.  
 TITLE The structure of V alpha and J alpha segments in the mouse  
 JOURNAL Nucleic Acids Res. 16 (23), 11355-11364 (1988)  
 MEDLINE 89083566  
 PUBMED 2849763

FEATURES  
 source  
 1..66  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 1..66  
 CDS  
 1..66  
 /partial  
 /codon\_start=1  
 /product="T-cell receptor alpha chain"  
 /protein\_id="AAA40195.1"  
 /db\_xref="GI:201215"  
 /translation="TTASLGKIQFGTGVVTPDI"  
 BASE COUNT 18 a 16 c 19 g 13 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 4.86e+03 Length: 66  
 Score: 49.00 Matches: 8  
 Percent Similarity: 80.00% Conservative: 0  
 Best Local Similarity: 80.00% Mismatches: 2  
 Query Match: 11.09% Indels: 0  
 DB: 10 Gaps: 0

US-09-599-087b-5 (1-81) x MUSTCAYJ (1-66)

OY 63 CysLysleuGluproGluProargLeuTTP 72  
 |||||  
 Db 23 TGCAGTTTGACAGACAGACCAGGTGTGG 52

RESULT 6  
 RNTCRAJ39 70 bp mRNA linear ROD 06-JAN-1998  
 LOCUS  
 DEFINITION R. norvegicus TCR alpha joining region, clone library VAB82F39.  
 ACCESSION Y09178  
 VERSION Y09178.1 GI:1669729  
 KEYWORDS TCR alpha joining region.  
 SOURCE Rattus norvegicus.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Rattus.  
 REFERENCES  
 1 (bases 1 to 70)  
 Torres-Nagel, N., Deutschlander, A., Hermann, T., Arden, B. and Hunig, T.  
 Control of TCR V alpha-mediated positive repertoire selection and alleoreactivity by differential J alpha usage and CDR3 alpha composition  
 Int. Immunol. 9 (10), 1441-1452 (1997)  
 JOURNAL 98013626  
 MEDLINE 9352349  
 PUBMED 2 (bases 1 to 70)  
 REFERENCES  
 Torres-Nagel, N.E.  
 Direct Submission  
 Submitted (31-OCT-1996) N.E. Torres-Nagel, Institut of Virology and Immunobiology, Department of Immunology, Versbacherstr. 7, 97078 Wuerzburg, FRG

FEATURES  
 source  
 1..70  
 /organism="Rattus norvegicus"  
 /strain="LEW.1F"  
 /db\_xref="taxon:10116"  
 /haplotype="Rn1f"  
 /cell\_type="peripheral T cell"  
 /clone\_id="VAB82F39"  
 /dev\_stage="adult"  
 /rearranged  
 1..70  
 /gene="TCRAJ"  
 <1..>70  
 /gene="TCRAJ"  
 /codon\_start=1  
 /protein\_id="CAA70377.1"  
 /db\_xref="GI:1785816"  
 /translation="AVETATGKIQFGTGVVTP"  
 <1..>70  
 /gene="TCRAJ"  
 /note="CDR3 region and J alpha region"  
 /evidence="experimental"  
 BASE COUNT 16 a 13 c 24 g 17 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 5.18e+03 Length: 70  
 Score: 49.00 Matches: 8  
 Percent Similarity: 80.00% Conservative: 0  
 Best Local Similarity: 80.00% Mismatches: 2  
 Query Match: 11.09% Indels: 0  
 DB: 10 Gaps: 0

US-09-599-087b-5 (1-81) x RNTCRAJ39 (1-70)

OY 63 CysLysleuGluproGluProargLeuTTP 72  
 |||||  
 Db 32 TGCAGTTTGACAGACAGACCAGGTGTGG 61

RESULT 7  
 S75385S18 76 bp DNA linear PRI 02-JUN-2000  
 LOCUS S75385S18

```

DEFINITION collagen type VI alpha 1(VI) chain (exons 3-19 and exon/intron
            junctions) [human, Genomic, 76 nt, segment 18 of 18].
ACCESSION  S75420
VERSION    S75420.1 GI:241983
KEYWORDS
SEGMENT   18 of 18
SOURCE     Homo sapiens.
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 76)
            Salta, B., Wang, Y.M., Renkart, L., Zhang, R.Z., Pan, T.C., Timpl, R.
            and Chu, M.L.
            The exon organization of the triple-helical coding regions of the
            human alpha 1(VI) and alpha 2(VI) collagen genes is highly similar
            Genomics 11 (1), 145-153 (1991)
JOURNAL    92112205
MEDLINE    1765372
PUBMED     1765372
REMARK     Genbank staff at the National Library of Medicine created this
            entry (NCBI gi241983 75420) from the original journal article.
            This sequence comes from Figure 2.
            Map location: chromosome 21.
FEATURES   Location/Qualifiers
            source          1..76
                           /organism="Homo sapiens"
                           /db_xref="taxon:9606"
                           /chromosome="21"
            mRNA           join(S75385.1:30..74,S75388.1:31..57,S75390.1:31..57,
                           S75392.1:31..75,S75394.1:31..84,S75396.1:31..93,
                           S75398.1:31..93,S75400.1:31..84,S75402.1:31..93,
                           S75404.1:31..93,S75406.1:31..93,S75408.1:31..93,
                           S75410.1:31..93,S75412.1:31..81,S75414.1:31..66,
                           S75416.1:31..93,S75418.1:31..96,31..>66)
                           /product="collagen type VI alpha 1(VI) chain"
            CDS             join(S75385.1:30..74,S75388.1:31..57,S75390.1:31..57,
                           S75392.1:31..75,S75394.1:31..84,S75396.1:31..93,
                           S75398.1:31..93,S75400.1:31..84,S75402.1:31..93,
                           S75404.1:31..93,S75406.1:31..93,S75408.1:31..93,
                           S75410.1:31..93,S75412.1:31..81,S75414.1:31..66,
                           S75416.1:31..93,S75418.1:31..96,31..>66)
                           /gene="collagen type VI alpha 1(VI) chain"
                           /note="This sequence comes from Figure 2; conceptual
                           translation differs from the translation provided in the
                           manuscript"
                           /codon_start=1
                           /product="collagen type VI alpha 1(VI) chain"
                           /protein_id="AAB20835.2"
                           /db_xref="GI:8176524"
                           /translation="GREDLGPVGIQGMKKGSGRSGKSGPKYKGEKGRKIDGV
                           DGVEKEMKPELPCKKSGPFGDIQGPFGPDGAGLKKKEKPEPADGAGRPGRAR*
                           GPGSDGEGPAGPPGPGGKSGAGDEGNPGPDGAPGEGRGRGTGTPGPGPDG
                           EAGPGDGGREGPVGPDGPEAGPIGPKYGRGSGGAGGAGGAPGPDGPGI
                           MGERGEDGPDGNGTEGFGPGPGYGRNGAPGINTKGYPLKDEGAGADGDDNNDI
                           APRGVKAKGVRGPEGPGGPGPRGPGPRGP"
            exon            31..>66
                           /number=20
BASE COUNT  15 a      26 c      24 g      11 t
ORIGIN
>>>
Alignment Scores:
Pred. No.: 7.04e+03 Length: 76
Score: 48.00 Matches: 9
Percent Similarity: 59.09% Conservative: 4
Best Local Similarity: 40.91% Mismatches: 9
Query Match: 10.86% Indels: 0
Gaps: 0
US-09-599-087b-5 (1-81) x S75385S18 (1-76)
QY 15 LeuCySPheserIlePheSerThrcIuclYlYsaRfargProAlaYsAlatrpsercyl 34
Db 6 CTCTCTACTCTCGTTTCTCGGACAGGAGCCGCCAGGACACCAAGACCGCTGGGCGCA 65

```

```

QY 35 ArgArg 36
Db 66 CGTAAG 71
RESULT 8
LOCUS      AX161822
DEFINITION Sequence 5150 from Patent WO0140521.
ACCESSION  AX161822
VERSION    AX161822.1 GI:14543153
KEYWORDS
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 51)
            Shinkets, R.A. and Leach, M.
            Nucleic acids containing single nucleotide polymorphisms and
            methods of use thereof
            Patent: WO 0140521-A 5150 07-JUN-2001;
            Curagen Corporation (US)
FEATURES   Location/Qualifiers
            source          1..51
                           /organism="Homo sapiens"
                           /db_xref="taxon:9606"
            misc_feature    26
                           /note="2 of 2 allelic variants (5149 is other entry)"
                           accession number cg43984006"
BASE COUNT  14 a      13 c      13 g      11 t
ORIGIN
>>>
Alignment Scores:
Pred. No.: 5.7e+03 Length: 51
Score: 47.00 Matches: 8
Percent Similarity: 72.73% Conservative: 0
Best Local Similarity: 72.73% Mismatches: 3
Query Match: 10.63% Indels: 0
Gaps: 0
US-09-599-087b-5 (1-81) x AX161822 (1-51)
QY 39 LeuCySPheserIlePheSerThrcIuclYlYsaRfargProAlaYsAlatrpsercyl 49
Db 19 TTGTCTGTGTGACGACGACCGTCGACGATTC 51
RESULT 9
LOCUS      AR077574
DEFINITION Sequence 289 from patent US 5962255.
ACCESSION  AR077574
VERSION    AR077574.1 GI:10004320
KEYWORDS
SOURCE     Unknown.
ORGANISM   Unknown.
            Unclassified.
            Griffiths, A.David., Williams, S.Cameron., Waterhouse, P.Michael.,
            Smith, A., John, Hammond.
            Smith, A., Winter, G., Paul., Johnson, K.Stuart. and
            Patent: US 5962255-A 289 05-OCT-1999;
            Methods for producing recombinant vectors
            Location/Qualifiers
            source          1..48
                           /organism="unknown"
BASE COUNT  8 a      18 c      11 g      11 t
ORIGIN
>>>
Alignment Scores:
Pred. No.: 5.95e+03 Length: 48
Score: 46.50 Matches: 9
Percent Similarity: 66.67% Conservative: 3
Best Local Similarity: 50.00% Mismatches: 3
Query Match: 10.52% Indels: 3

```





GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 8, 2003, 22:13:48 ; Search time 1348 seconds  
(without alignments)

973.171 Million cell updates/sec

Title: US-09-599-087b-5

Perfect score: 442

Sequence: 1 MRLVLSLCLILLCFSIF.....PCKLEPPRLMVPGLAPGV 81

## Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 227956

Minimum DB seq length: 25  
Maximum DB seq length: 80

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

## Command line parameters:

-MODEL=frame\_plus\_p2n.model -DEV=xlp  
-Q/cgml2.1/USPTO.spool/US09599087/runat\_07052003.113852.7677/app\_query.fasta.1.263  
-DB=EST -QFMT=fastap -SUFFIX=est -MIMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pco -NORM=ext -HEA\_SIZE=500 -MINLEN=25 -MAXLEN=80  
-USER=US09599087.qcgnl.1.1525\_xrunat\_07052003.113852.7677 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_othr:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed; and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	50.5	11.4	60	17	AZ918344
C 2	50.5	11.4	69	17	BH639950
C 3	49	11.1	72	14	BQ394194
C 4	49	11.1	76	17	AZ918366
C 5	49	11.1	79	9	A1877052
C 6	48	10.9	80	10	AV676540
C 7	47	10.6	67	17	AZ605110
C 8	47	10.6	74	9	AA739046
C 9	47	10.6	73	10	AM600148
C 10	47	10.6	74	13	BM283913
C 11	46.5	10.5	79	9	A1468972
C 12	46	10.4	50	9	AU105245
C 13	46	10.4	70	14	BQ092063
C 14	46	10.4	70	14	BQ092271
C 15	46	10.4	73	14	H22549
C 16	46	10.4	77	9	AA049559
C 17	45.5	10.3	77	14	BQ666094
C 18	45.5	10.3	77	14	BQ811778
C 19	45	10.2	67	13	BM518515
C 20	45	10.2	70	17	AZ992067
C 21	45	10.2	72	14	BQ757145
C 22	45	10.2	72	17	AZ493456
C 23	45	10.2	72	17	AZ921678
C 24	45	10.2	73	9	AA790708
C 25	45	10.2	79	9	AA498761
C 26	45	10.2	80	9	AU257386
C 27	44	10.0	74	13	BM097862
C 28	44	10.0	76	9	AU244358
C 29	44	10.0	78	6	BQ241316
C 30	43.5	9.8	67	17	BH620911
C 31	43.5	9.8	67	17	AZ808082
C 32	43.5	9.8	72	14	R33083
C 33	43	9.7	40	9	A1589916
C 34	43	9.7	40	17	TA176612P
C 35	43	9.7	51	10	AA232755
C 36	43	9.7	58	9	AA948249
C 37	43	9.7	58	17	AZ835896
C 38	43	9.7	62	17	FR0012583
C 39	43	9.7	64	10	AM626440
C 40	43	9.7	64	10	AM626725
C 41	43	9.7	64	12	BF118532
C 42	43	9.7	70	9	AU258504
C 43	43	9.7	72	10	AV911900
C 44	43	9.7	75	9	AL643544
C 45	43	9.7	75	13	BI753134

## ALIGNMENTS

RESULT 1  
AZ918344/c  
LOCUS  
DEFINITION  
1006004B02.xl 1006 - Rescuenu G1d G Zea mays genomic, DNA  
60 bp  
DNA  
linear  
GSS 17-DEC-2001  
sequence.  
AZ918344  
ACCESSION  
AZ918344.1 GI:13387628  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Zea mays.  
Zea mays.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE  
1 (bases 1 to 60)  
Walbot,Y.

TITLE Maize genomic sequences found using engineered Rescuemu transposon  
JOURNAL Unpublished (2001)  
COMMENT Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221

Email: walbot@stanford.edu  
Possible ligation site of ends cut by 2 different endonucleases.  
Reverse complemented post-ligation sequence from source sequence.  
Plate: 1006004 row: 36  
Class: transposon-tagged.

# FEATURES

## source

1..60  
/organism="Zea mays"  
/cultivar="mixed background W23/A188/B73"  
/db\_xref="taxon:4577"  
/clone\_lib="1006 - Rescuemu Grid G"  
/tissue\_type="leaf"  
/dev\_stage="adult"  
/lab\_host="DH10B"

/note="Organ: leaf; Vector: Rescuemu (engineered from pBluescript backbone); Site\_1: BamHI; Site\_2: BglII; Rescuemu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on Rescuemu, go to the web site 'www.zmdb.lastate.edu' and follow the links for 'Rescuemu.' Grid G was grown at Stanford in 2000. DNA was extracted from leaf punches, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."

BASE COUNT 5 a 17 c 26 g 12 t  
ORIGIN

## Alignment Scores:

Pred. No.: 9.11e+03 Length: 60  
Score: 50.50 Matches: 10  
Percent Similarity: 78.57% Conservative: 1  
Best Local Similarity: 71.43% Mismatches: 2  
Query Match: 11.43% Indels: 1  
DB: 17 Gaps: 1

US-09-599-087b-5 (1-81) x A2918344 (1-60)

Qy 34 G1YARgArGThrArgLeuCySHsArgValProSerPro 47

Db 43 GGCACCGACGACGCGCTCATCTGCACCGG---CCACACGCCA 5

RESULT 2  
BH639950/c 69 bp DNA linear GSS 14-FEB-2002  
LOCUS 1008032807.2EL\_y1 1008 - Rescuemu Grid I Zea mays genomic, DNA  
DEFINITION  
sequence.  
ACCESSION BH639950  
VERSION BH639950.1 GI:18665756  
KEYWORDS  
SOURCE Zea mays.  
ORGANISM Zea mays.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 69)

REFERENCE  
AUTHORS Walbot V.  
TITLE Maize genomic sequences found using engineered Rescuemu transposon  
JOURNAL Unpublished (2001)  
COMMENT Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227

Fax: 650 725 8221  
Email: walbot@stanford.edu  
Possible ligation site of ends cut by 2 different endonucleases.  
Reverse complemented post-ligation sequence from source sequence.  
Plate: 1008032 row: 22  
Class: transposon-tagged.

# FEATURES

## source

1..69  
/organism="Zea mays"  
/cultivar="mixed background W23/A188/B73"  
/db\_xref="taxon:4577"  
/clone\_lib="1008 - Rescuemu Grid I"  
/tissue\_type="leaf"  
/dev\_stage="adult"  
/lab\_host="DH10B"

/note="Organ: leaf; Vector: Rescuemu (engineered from pBluescript backbone); Site\_1: BamHI; Site\_2: BglII; Rescuemu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on Rescuemu, go to the web site 'www.zmdb.lastate.edu' and follow the links for 'Rescuemu.' Grid I was grown at Berkeley in 2001. DNA was extracted from leaf punches, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."

BASE COUNT 7 a 30 c 26 g 6 t  
ORIGIN

## Alignment Scores:

Pred. No.: 1.11e+04 Length: 69  
Score: 50.50 Matches: 10  
Percent Similarity: 66.67% Conservative: 2  
Best Local Similarity: 55.56% Mismatches: 5  
Query Match: 11.43% Indels: 1  
DB: 17 Gaps: 1

US-09-599-087b-5 (1-81) x BH639950 (1-69)

Qy 25 LysARgArGProAlaLysAlaTrpSerGlyARgArGThrArgLeuCySHs 42

Db 62 CGCCGAGGCGCGCGCGCGGTGGCGC---CGCGTGACCGGAGTGCTCCAC 12

RESULT 3  
BQ394194/c 72 bp mRNA linear EST 22-MAY-2002  
LOCUS NISC\_ng07g11.x1 NICHD XCC Emb6 Silurana tropicalis cDNA clone  
DEFINITION  
IMAGE:5382740 3', mRNA sequence.  
ACCESSION BQ394194  
VERSION BQ394194.1 GI:21081881  
KEYWORDS  
SOURCE EST.  
ORGANISM western clawed frog.  
Silurana tropicalis  
Eukaryota; Metazoa; Chordata; Granulata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
Xenopodinae; Silurana.

REFERENCE  
AUTHORS NIH-XCG  
TITLE NIH-XCG http://image.llnl.gov/image/html/xenopuslib.info.shtml.  
JOURNAL National Institute of Child Health and Human Development, National  
Cancer Institute, Xenopus Gene Collection  
COMMENT Unpublished (2002)  
Contact: Robert Strausberg, Ph.D.  
Email: cgeabs-remail.nih.gov  
CDNA Library Preparation:  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC)  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
info@image.llnl.gov  
Plate: LHAM1974 row: N column: 21  
Seq primer: -21M13 forward primer (AB1).





rounds of normalization, and was constructed by Bento Soares and M. Fatima Bernaldo."

BASE COUNT 15 a 23 c 25 g 16 t

## Alignment Scores:

Pred. No.: 1.95e+04 Length: 79  
Score: 49.00 Matches: 10  
Percent Similarity: 47.83% Conservative: 12  
Best Local Similarity: 43.48% Mismatches: 12  
Query Match: 11.09% Indels: 0  
DB: 9 Gaps: 0

US-09-599-087b-5 (1-81) x A1877052 (1-79)

OY 40 CysCysHsArgValProSerProAsnSerThrAsnLeuLysGlyHisHsValArgLeu 59  
DB 78 TGCCAGAGCAGTCTCTCTAAGCTCCGGAATCCATCCAGCGAGAGCAGAGC 19

OY 60 CysLysPro 62  
DB 18 TGTCTCTCC 10

RESULT 6 AV676540 80 bp mRNA linear EST 05-OCT-2000  
AV676540  
LOCUS AV676540  
DEFINITION cDNA clone rclb1e23 3', mRNA sequence.

ACCESSION AV676540  
VERSION AV676540.1 GI:10114539  
KEYWORDS EST.  
SOURCE Clona intestinalis.  
ORGANISM Clona intestinalis.

REFERENCE Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;  
Phlebobranchia; Cloniidae; Clona.

AUTHORS 1 (bases 1 to 80)  
TITLE Satoh, N., Satou, Y., Kohara, Y. and Shin-I, T.  
JOURNAL Expressed genes in Clona intestinalis  
COMMENT Unpublished (2000)  
Contact: Nori Satoh  
Department of Zoology  
Kyoto University  
Sakyo-ku, Kyoto, Kyoto 606-8502, Japan  
Tel: 81-75-753-4081  
Fax: 81-75-705-1113  
Email: satoh@ascidian.zool.kyoto-u.ac.jp.

## FEATURES

Location/Qualifiers  
1..80  
/organism="Clona intestinalis"  
/db\_xref="taxon:7719"  
/clone="rclb1e23"  
/clone\_1b="Nori Satoh unpublished cDNA library"  
/issue\_type="whole animal"  
/dev\_stage="tailbud"  
/note="Vector: pBluescript SK"

BASE COUNT 20 a 19 c 15 g 26 t  
ORIGIN

## Alignment Scores:

Pred. No.: 2.53e+04 Length: 80  
Score: 48.00 Matches: 8  
Percent Similarity: 68.42% Conservative: 5  
Best Local Similarity: 42.11% Mismatches: 4  
Query Match: 10.86% Indels: 2  
DB: 10 Gaps: 1

US-09-599-087b-5 (1-81) x AV676540 (1-80)

OY 32 TrpSerGlyArgGlyThrArgLeuGlyCysHisArg-----ValProSerProAsn 48  
DB 15 TGGGAGGTAAGATGCTCTGTGTCTGTCATCCAAATATGTTATTCAGGCCAGAT 71  
RESULT 7

AZ605110/c 67 bp DNA linear GSS 13-DEC-2000  
LOCUS AZ605110  
DEFINITION 1M0426115F Mouse 10kb plasmid UGCCIM library Mus musculus genomic  
clone U06C1M0426115 F, DNA sequence.

ACCESSION AZ605110  
VERSION AZ605110.1 GI:11727300  
KEYWORDS GSS.  
SOURCE house mouse.  
ORGANISM Mus musculus

## REFERENCE

AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamli, C.,  
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,  
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.  
and Wright, D., Weiss, R.  
TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0426 row: 1 column: 15  
Seq primer: CGTGTGAACGACGCGCACT  
Class: plasmid ends  
High quality sequence stop: 67.

## FEATURES

Location/Qualifiers

1..67  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="U06C1M0426115"  
/clone\_1b="Mouse 10kb plasmid UGCCIM library"  
/sex="Male"  
/lab\_host="E. Coli strain XL10-Gold, T1-resistant, F-"  
/note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA was hydronically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (9114732114191AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 15 a 17 c 22 g 13 t  
ORIGIN

## Alignment Scores:

Pred. No.: 2.49e+04 Length: 67  
Score: 47.00 Matches: 8  
Percent Similarity: 70.59% Conservative: 4  
Best Local Similarity: 47.06% Mismatches: 5  
Query Match: 10.63% Indels: 0  
DB: 17 Gaps: 0

US-09-599-087b-5 (1-81) x AZ605110 (1-67)

QY 44 ValProserProasnsrThraenLeuylsGlyHshValargLeucys 60  
 ||||| ||||| :||||| ||| ||| :|||||  
 Db 66 GTCCACACCCCAACCCCTCTCTGTGTAGGGGTCAATCAAGCTTTGT 16

RESULT 8  
 AM600148 73 bp mRNA linear EST 22-MAR-2000  
 LOCUS SMLC4K10E07SK Brugia malayi L4 CDNA (SAM99MLM-Bml4) Brugia malayi  
 DEFINITION CDNA clone SMLC4K10E07 5', mRNA sequence.

ACCESSION AM600148  
 VERSION AM600148.1 GI:7287661  
 KEYWORDS EST.  
 SOURCE Brugia malayi.  
 ORGANISM Brugia malayi.  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;  
 Onchocercidae; Brugia.

REFERENCE  
 AUTHORS Williams, S.A.  
 TITLE Genes expressed in fourth stage larvae of Brugia malayi  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Steven A. Williams  
 Molecular Parasitology  
 Smith College Department of Biological Sciences  
 Department of Biological Sciences, Clark Science Center, Smith  
 College, Northampton, MA, 01063, USA  
 Tel: 4135853826  
 Fax: 4135853786  
 Email: genomesmith.edu  
 Seq primer: pbluescript SK.  
 Location/Qualifiers  
 1..73  
 /organism="Brugia malayi"  
 /db\_xref="taxon:6279"  
 /clone="SMLC4K10E07"  
 /clone\_1lb="Brugia malayi L4 CDNA (SAM99MLM-Bml4)"  
 /dev\_stage="larval stage four"  
 /lab\_host="XLI-Blue MRF"  
 /note="Vector: Lambda Uni-ZAP XR; Site\_1: Eco RI; Site\_2:  
 Xho I; Lymphatic filarial nematode parasite of humans.  
 mRNA was prepared from L4s isolated from the peritoneal  
 cavity of jirds and converted to double-stranded cDNA  
 using reverse transcriptase and oligo(dT) followed by  
 RNase H and DNA pol I. The library has 2.7 x 10<sup>5</sup>  
 independent recombinants and the average insert size is  
 approx. 1050bp. The library was constructed by Michelle  
 Lizotte-Waniewski. The library is available from Dr. S.A.  
 Williams, email: genomesmith.edu."

BASE COUNT 21 a 11 c 10 g 31 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 2.82e+04 Length: 73  
 Score: 47.00 Matches: 10  
 Percent Similarity: 70.59% Conservative: 2  
 Best Local Similarity: 58.82% Mismatches: 5  
 Query Match: 10.63% Indels: 0  
 DB: 10 Gaps: 0

US-09-599-087b-5 (1-81) x AM600148 (1-73)

QY 6 LeuSerLeuLeuCySileLeuLeuCySphSerIlePheSerThr 22  
 ||||| :||| ||||| ||||| :||| |||  
 Db 3 CATACATCACTGTCTTTGTCATTTGCTTTTTCATTTTGAACCA 53

RESULT 9  
 AA739046 74 bp mRNA linear EST 14-JAN-1998  
 LOCUS AA739046/c  
 DEFINITION v66d04.v1 Stratiogene mouse skin (#937313) Mus musculus cDNA clone  
 IMAGE:1227367 5', similar to gp:M73436 Mouse ribosomal protein S4  
 (MOUSE);, mRNA sequence.

ACCESSION AA739046  
 VERSION AA739046.1 GI:2775852  
 KEYWORDS EST.

SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 74)  
 REFERENCE  
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
 Geisler, S., Kucaba, T., Lacey, M., Le, M., Martin, J., Morris, M.,  
 Schellinger, B., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.  
 TITLE The WashU-HMNI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT Contact: Marra M/Mouse EST Project  
 WashU-HMNI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:652959  
 Seq primer: -28m13 rev1 ET from Amersham.

FEATURES  
 source  
 1..74  
 /organism="Mus musculus"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:1227367"  
 /clone\_1lb="Stratiogene mouse skin (#937313)"  
 /sex="females"  
 /tissue\_type="whole skin"  
 /dev\_stage="11 weeks old"  
 /lab\_host="SOLR (kanamycin resistant)"  
 /note="Organ: skin; Vector: pbluescript SK-; Site\_1: EcoRI  
 ; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo  
 dt. Whole skin from 11 week old C57BL/6 female mice.  
 Average insert size: 1.0 kb. Uni-ZAP XR Vector; -5'  
 adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor  
 sequence: 5' CTCGAGTCTTTTCTTTTCTTTTCTTTT 3'."

BASE COUNT 19 a 24 c 15 g 16 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 2.88e+04 Length: 74  
 Score: 47.00 Matches: 8  
 Percent Similarity: 83.33% Conservative: 2  
 Best Local Similarity: 66.67% Mismatches: 0  
 Query Match: 10.63% Indels: 0  
 DB: 9 Gaps: 0

US-09-599-087b-5 (1-81) x AA739046 (1-74)

QY 47 ProAsnsrThraenLeuylsGlyHshValargLeucys 58  
 ||||| ||||| :||||| ||| ||| :|||||  
 Db 54 CCGAATAGACGACGATGCGTCAACGACGATCGG 19

RESULT 10  
 BM283913 74 bp mRNA linear EST 28-DEC-2001  
 LOCUS BM283913/c  
 DEFINITION k135b04.v1 Ascaris suum female gonad G2 SL1 TOPO v1 Murphy  
 Chiapelli1 McCarter Ascaris suum cDNA 5', mRNA sequence.

ACCESSION BM283913  
 VERSION BM283913.1 GI:17992955  
 KEYWORDS EST.  
 SOURCE pig roundworm.  
 ORGANISM Ascaris suum  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea  
 ; Ascarididae; Ascaris.  
 1 (bases 1 to 74)  
 REFERENCE  
 AUTHORS McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,  
 Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,  
 Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagarelis, V., R.





```

/organism="Homo sapiens"
/db_xref="GDB:3835770"
/db_xref="taxon:9606"
/clone="IMAGE:173735"
/clone_11b="Scores adult brain N25HB55"
/sex="Male"
/dev_stage="55-year old"
/lab_host="DH10B (ampicillin resistant)"
/notes="organ: brain; vector: pT73D (Pharmacia) with a
modified polylinker; site_1: Not I; site_2: Eco RI; 1st
stratified cDNA was primed with a Not I - oligo(dT) primer
[5'-GTTTTCACATGTGAAGTGGGAGCGCGCGCGCTTTTTTTTTTTTTT
3'],

```

Search completed: May 8, 2003, 23:05:56  
Job time : 1349 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 21:43:48 ; Search time 30 Seconds

(without alignments)  
556.327 Million cell updates/sec

Title: US-09-599-087b-5

Perfect score: 442  
Sequence: 1 MRLVLSLLCLILLCFSIF.....PCKLPEPRHLYVPGALPOV 81

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 69166

Minimum DB seq length: 25

Maximum DB seq length: 80

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

SPREMBL\_21: \*  
1: sp\_archaea: \*  
2: sp\_bacteria: \*  
3: sp\_fungi: \*  
4: sp\_human: \*  
5: sp\_invertebrate: \*  
6: sp\_mammal: \*  
7: sp\_mhc: \*  
8: sp\_organelle: \*  
9: sp\_phage: \*  
10: sp\_plant: \*  
11: sp\_rodent: \*  
12: sp\_virus: \*  
13: sp\_vertebrate: \*  
14: sp\_unclassified: \*  
15: sp\_virus: \*  
16: sp\_bacteriophage: \*  
17: sp\_archaeal: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	56.5	12.8	57	4	Q96HV2
2	56.5	12.8	78	10	Q8VZQ7
3	56.5	12.7	69	5	Q9BXX4
4	54.5	12.3	75	16	Q9KIL9
5	53	12.0	64	16	Q8UE52
6	53	12.0	78	12	Q8VAV9
7	52.5	11.9	69	12	Q8ON69
8	51	11.5	51	16	Q985E2
9	51	11.5	79	16	Q9KKN6
10	50	11.3	42	8	Q32389
11	50	11.3	59	12	Q9Q6Q2
12	50	11.3	76	12	Q9JXH1
13	49.5	11.2	80	11	Q9JHY4
14	48.5	11.0	71	5	Q9VU66
15	48	10.9	61	11	Q61910
16	48	10.9	69	2	Q9ZC27

17	48	10.9	79	5	017707	017707 caenorhabdi
18	47.5	10.7	69	16	Q9EWV8	Q9EWV8 streptomyces
19	47.5	10.7	77	6	Q28593	Q28593 ovis aries
20	47.5	10.7	77	6	Q95M00	Q95M00 ovis aries
21	47.5	10.7	77	16	Q8XCT7	Q8XCT7 escherichia
22	47	10.6	45	2	Q52365	Q52365 escherichia
23	47	10.6	64	12	Q8U2B1	Q8U2B1 garlic late
24	47	10.6	77	16	Q932P0	Q932P0 staphylococ
25	46.5	10.5	55	4	Q8TBY0	Q8TBY0 homo sapien
26	46.5	10.5	72	6	Q9N072	Q9N072 macaca fasc
27	46	10.4	36	4	Q9BR19	Q9BR19 homo sapien
28	46	10.4	39	5	Q9XG8	Q9XG8 plasmodium
29	46	10.4	57	10	Q851R9	Q851R9 oryza sativ
30	46	10.4	70	10	Q9FHN5	Q9FHN5 arabidopsis
31	46	10.4	71	5	Q95N85	Q95N85 drosophila
32	46	10.4	80	8	Q35501	Q35501 paracentrot
33	45.5	10.3	48	4	Q9P1E0	Q9P1E0 homo sapien
34	45.5	10.3	60	5	Q9BJT5	Q9BJT5 plasmodium
35	45.5	10.3	61	10	Q94DR7	Q94DR7 oryza sativ
36	45.5	10.3	66	10	Q943X6	Q943X6 oryza sativ
37	45.5	10.3	77	16	Q94R85	Q94R85 vibrrio chol
38	45	10.2	32	8	Q36494	Q36494 farfantepen
39	45	10.2	38	5	P91965	P91965 penaeus van
40	45	10.2	40	8	Q9TGM1	Q9TGM1 homo sapien
41	45	10.2	44	11	Q04056	Q04056 rattus norv
42	45	10.2	49	2	Q50145	Q50145 mycobacteri
43	45	10.2	51	17	Q8TIN2	Q8TIN2 methanosarc
44	45	10.2	54	16	Q92P07	Q92P07 rhizobium m
45	45	10.2	56	13	Q42594	Q42594 xenopus lae

## ALIGNMENTS

### RESULT 1

Q96HV2 ID Q96HV2 PRELIMINARY: PRT: 57 AA.  
AC Q96HV2;  
DT 01-DEC-2001 (TREMREL. 19, Created)  
DT 01-DEC-2001 (TREMREL. 19, Last sequence update)  
DT 01-DEC-2001 (TREMREL. 19, Last annotation update)  
DE Similar to glutamate receptor, metabotropic 5.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PLACENTA;  
RA Strausberg R.;  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC008050; AAH08050.1;  
KW Receptor.  
SQ SEQUENCE 57 AA: 5734 MW: DDAF557866DB753A CRC64;

Query Match 12.8%; Score 56.5; DB 4; Length 57;  
Best Local Similarity 33.3%; Pred. No. 8.5;  
Matches 18; Conservative 5; Mismatches 16; Indels 15; Gaps 3;

QY 23 EGRKRPKAWGGRTRLCCHRVSPNSNLMGHVRLCKPCKLPEPRHLYVPG 76  
Db 13 KGGGPKAWMEGPTN-----SSPNEAN-----QOCEGTQGP--LEAPG 51

### RESULT 2

Q8VZQ7 ID Q8VZQ7 PRELIMINARY: PRT: 78 AA.  
AC Q8VZQ7;  
DT 01-MAR-2002 (TREMREL. 20, Created)  
DT 01-MAR-2002 (TREMREL. 20, Last sequence update)  
DT 01-JUN-2002 (TREMREL. 21, Last annotation update)  
DE Putative defensin Ampi protein.  
GN ATIG19610.

OS Arabidopsis thaliana (Mouse-ear cress). Embryophyta; Tracheophyta;  
 OC Eukaryota; Viridiplantae; Streptophyta; Eudicotyledons; core eudicots; Rosidae;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 RN NCBI\_Taxid=3702;  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Banh J., Banno F., Chang E., Dale J.M., Goldsmith A.D.,  
 RA Lee J.M., Onodera C.S., Quach H.L., Tang C., Tortum M., Wu H.C.,  
 RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,  
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai J., Kim C., Kosemura E., Lam B., Lin J., Meyers M.C., Miranda M.,  
 RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Setou M., Seki M.,  
 RA Shinn P., Southwick A., Shinzaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.,  
 RT "Full length cDNA of gene Atg191610 (GI:15223595)."  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY063933; AAL36289.1;  
 DR InterPro: IPR002118; Gamma-thionin.  
 DR InterPro: IPR003614; Knott.  
 DR Pfam: PF00304; Gamma-thionin. 1.  
 DR ProDom: PD002594; Gamma-thionin. 1.  
 DR SMART: SM00505; Knott. 1.  
 DR PROSITE: PS00940; GAMMA\_THIONIN; UNKNOWN.1.  
 SQ SEQUENCE 78 AA; 8840 MW; A5B5DD2830A6545 CRC64;  
 Query Match 12.8%; Score 56.5; DB 10; Length 78;  
 Best Local Similarity 36.5%; Pred. No. 11;  
 Matches 19; Conservative 5; Mismatches 13; Indels 15; Gaps 4;  
 Oy 7 SSLLCILLLEFIF-----STEEK--RRPAKAWSG--RRPLG---CHR 43  
 Db 3 SSTYLMFLCLSTFLINASTEMAAVEGRICERKRTWTGCGNTRGCDSCQR 54  
 RESULT 3  
 09BKX4 PRELIMINARY; PRT; 69 AA.  
 AC 09BKX4;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Hypothetical 7.7 kDa protein.  
 OS Y22D7AR.10.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 RN NCBI\_Taxid=6239;  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-BRISTOL N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode C. elegans: a platform for  
 RT investigating biology. The C. elegans Sequencing Consortium.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-BRISTOL N2;  
 RA Madsen C.;  
 RT "The sequence of C. elegans cosmid Y22D7AR.";  
 RL submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-BRISTOL N2;  
 RA Waterston R.;  
 RT "Direct Submission";  
 RL EMBL: AC084154; AAR29877.1;  
 KW Hypothetical protein.  
 SQ SEQUENCE 69 AA; 7682 MW; 9DD66CD46976CF7E CRC64;  
 Query Match 12.7%; Score 56; DB 5; Length 69;

Best Local Similarity 29.1%; Pred. No. 12;  
 Matches 23; Conservative 8; Mismatches 30; Indels 18; Gaps 6;  
 Oy 1 MRLVLSSLLCILLCSIFSTEGKRPA--KAMSGRRRLCCHRVSPNSTNLKGHV 57.  
 Db 1 MRLFISSL-LILLCKRLGDTCDYGCFFKSISSGR---CC-----PDGWEFGGR-- 49  
 Oy 58 RLCKPKCLEPE-RLWVVP 75  
 Db 50 ---APCPANPMDEWVCCP 65  
 RESULT 4  
 09KIL9 PRELIMINARY; PRT; 75 AA.  
 AC 09KIL9;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE Hypothetical protein NMB0092.  
 GN NMB0092.  
 OS Neisseria meningitidis (serogroup B).  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 RN NCBI\_Taxid=491;  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-MC58 / SEROGROUP B;  
 RX MEDLINE=20175755; PubMed=10710307;  
 RA Tetelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
 RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,  
 RA Nelson W.C., Gwinn M.L., Deboy R., Peterson J.D., Hickey E.K.,  
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
 RA Mason T., Ciecko A., Parney D.S., Blair E., Cifton H., Clark E.B.,  
 RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,  
 RA Gill J., Scarlato V., Maignani V., Plaza M., Grandi G., Sun L.,  
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappelli R., Venter J.C.;  
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain  
 RT MC58.";  
 RL Science 287:1809-1815(2000).  
 DR EMBL: AE002368; AAF40554.1;  
 DR TIGR: NMB0092;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 75 AA; 8504 MW; B60F7F15CC8ACFB8 CRC64;  
 Query Match 12.3%; Score 54.5; DB 16; Length 75;  
 Best Local Similarity 29.7%; Pred. No. 20;  
 Matches 22; Conservative 8; Mismatches 25; Indels 19; Gaps 4;  
 Oy 1 MRLVLSSLLCILLCSIFSTEGKRPAKAMSGRRRLCCH-RVSPNSTNLKGH 55  
 Db 2 VRFVLSFTLLINCSISACNSHF-----TGNINPLGTHNKVNPNCANSANS 49  
 Oy 56 HVRLLCKPKCLEPEP 69  
 Db 50 HIR--QPSKNYDP 61  
 RESULT 5  
 08UE52 PRELIMINARY; PRT; 64 AA.  
 AC 08UE52;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Hypothetical protein Atu1909.  
 GN Atu1909.  
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Rhizobium.  
 RN NCBI\_Taxid=176299;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21608550; PubMed=11743193;



RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,  
 RA Chen Y., Paulsen I.T., Eisen J.A., Kap P.D., Bovee D., Strick,  
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,  
 RA Kutyavin T., Levy R., Li M.-J., McEllland E., Palmeri A.,  
 RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,  
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
 RA Nester E.W.,  
 RT "The genome of the natural genetic engineer *Agrobacterium tumefaciens*  
 RT C58.";  
 RL Science 294:2317-2323(2001).  
 DR EMBL: AE009145; AAL42905.1;  
 KM Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 64 AA; 7544 MW; C15A5CA8687C42BE CRC64;  
 Query Match 12.0%; Score 53; DB 16; Length 64;  
 Best Local Similarity 28.0%; Pred. No. 26;  
 Matches 14; Conservative 6; Mismatches 16; Indels 14; Gaps 1;  
 QY 25 KRPAKAWSGRRRLCCHRPSPNSTNLGHHVRLCKPCKLEPEPLWVY 74  
 DB 17 KRRPAR-----PVKRTALSGVRLCKGMQWENDPREWMSI 52  
 RESULT 6  
 QY 08VAU9 PRELIMINARY; PRT; 78 AA.  
 AC 08VAU9;  
 DT 01-MAR-2002 (TREMBLrel. 20, Created)  
 DT 01-JUN-2002 (TREMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE WSV280 (WSSV335).  
 OS White spot syndrome virus (WSSV).  
 OC Viruses; unclassified viruses.  
 OX NCBI\_TaxID=92652;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21548311; PubMed=11689662;  
 RA Yang F., He J., Lin X., Li Q., Pan D., Zhang X., Xu X.,  
 RT "Complete genome sequence of the shrimp white spot bacilliform  
 RT virus";  
 RL J Virol. 75:11811-11820(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Yang F., He J., Lin X., Li Q., Pan D., Zhang X., Xu X.,  
 RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-TAIWAN;  
 RX MEDLINE=20517548; PubMed=11062040;  
 RA Tsai M.F., Yu H.T., Tzeng H.F., Liu J.H., Chou C.M., Huang C.J.,  
 RA Wang C.H., Lin J.Y., Kou G.H., Lo C.F.,  
 RT "Identification and characterization of a shrimp white spot syndrome  
 RT virus (WSSV) gene that encodes a novel chimeric polypeptide of  
 RT cellular-type thymidine kinase and thymidylate kinase";  
 RL Virology 277:100-110(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-TAIWAN;  
 RX MEDLINE=21844071; PubMed=11853398;  
 RA Chen L.L., Liu J.H., Huang C.J., Chou C.M., Chen S.M., Wang C.H.,  
 RA Lo C.F., Kou G.H.,  
 RT "Identification of a nucleocapsid protein (VP35) gene of shrimp white  
 RT spot syndrome virus and characterization of the motif important for  
 RT targeting VP35 to the nuclei of transsected insect cells";  
 RL Virology 293:44-53(2002).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-TAIWAN;  
 RT Lo C.-F., Kou G.-H.,  
 RL Submitted (Oct-2001) to the EMBL/Genbank/DBJ databases.

DR EMBL: AF332093; AAL33283.1;  
 DR EMBL: AF440570; AAL89203.1;  
 SQ SEQUENCE 78 AA; 8949 MW; B5800F9CCB754D0 CRC64;  
 Query Match 12.0%; Score 53; DB 12; Length 78;  
 Best Local Similarity 32.1%; Pred. No. 32;  
 Matches 18; Conservative 7; Mismatches 27; Indels 4; Gaps 2;  
 QY 2 RLIVSSLLCILLCSIFSTESCKRRPAKAWSGRRRLCCHRPSPNSTNLKGH 56  
 DB 14 RLIVFTSLSSRVSLIFKSSMLRORVWSLFRKALCTVQSGAP---MIAQH 66  
 RESULT 7  
 QY 08ON69 PRELIMINARY; PRT; 69 AA.  
 AC 08ON69;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE ESV-1-221.  
 OS Ectocarpus siliculosus virus.  
 OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Phaeovirus.  
 OX NCBI\_TaxID=37665;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ESV-1;  
 RA Delarogue N., Bothe G., Pohl T., Knippers R., Mueller D.G., Boland W.,  
 RT "The complete nucleotide sequence of the Ectocarpus siliculosus virus  
 RT genome";  
 RL Submitted (Mar-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF204951; AKI4635.1;  
 SQ SEQUENCE 69 AA; 8160 MW; AACBD15F312D639 CRC64;  
 Query Match 11.9%; Score 52.5; DB 12; Length 69;  
 Best Local Similarity 25.8%; Pred. No. 33;  
 Matches 16; Conservative 10; Mismatches 23; Indels 13; Gaps 3;  
 QY 19 IFSTES-----KRPAAKWSGR--RTRLCCHRPSPNSTNLKGHHVRLCKPCKL 65  
 DB 2 IFVNGILNTKTLRRHRLCHLMSNNLIPRRRIPCNRLOSRAVRSRGAHKKICIGXY 61  
 QY 66 EP 67  
 DB 62 RP 63  
 RESULT 8  
 QY 0985E2 PRELIMINARY; PRT; 51 AA.  
 AC 0985E2;  
 DT 01-OCT-2001 (TREMBLrel. 18, Created)  
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Hypothetical protein ms17717.  
 GN MS17717.  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Phyllobacteriaceae; Mesorhizobium.  
 OX NCBI\_TaxID=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MAF303099;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Watanabe A., Ideasa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
 RA Mochizuki T., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.,  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT Mesorhizobium loti";  
 RL DNA Res. 7:331-338(2000).  
 DR EMBL: AP003012; BAB54120.1; -

KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 51 AA; 5843 MW; 4934C9A7166468F7 CRC64;

Query Match 11.5%; Score 51; DB 16; Length 51;  
 Best Local Similarity 32.0%; Pred. No. 38;  
 Matches 16; Conservative 12; Mismatches 12; Indels 10; Gaps 5;

OY 8 SLICILL-...LCFSFSTEGKRRPAKAMSGRTRLCCHRVSPNSTNLK 53  
 DB 7 TIVCLMSLFGEFALIVAEGRSP-RSWD-EATRIG-RLACP-TNLR 50

RESULT 9

OY 09KKN6 PRELIMINARY; PRT; 79 AA.  
 AC 09KKN6;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Hypothetical protein VCA1066.  
 GN VCA1066.  
 OS Vibrio cholerae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=666;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;  
 RX MEDLINE=20406833; PubMed=10952301;  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gilm M.L.,  
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Diragol I., Sellers P.,  
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
 RA Fraser C.M.;  
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
 RT cholerae";  
 RL Nature 406:477-483(2000).  
 DR EMBL; AE004432; AAF96960.1; -  
 RT TIGR; VCA1066; -  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 79 AA; 8342 MW; 716F1729FE8C41C2 CRC64;

Query Match 11.5%; Score 51; DB 16; Length 79;  
 Best Local Similarity 42.9%; Pred. No. 58;  
 Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

OY 39 LCCHRVSPNSTNLKGHVRL 59  
 DB 50 LTCHLPATPSSLGIGSHVRL 70

RESULT 10

OY 032989 PRELIMINARY; PRT; 42 AA.  
 AC 032989;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE ORF43f.  
 OS Pinus thunbergii (Green pine) (Japanese black pine).  
 OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.  
 OX NCBI\_TaxID=3350;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92212283; PubMed=1557027;  
 RA Tsudzuki J., Nakashima K., Tsudzuki T., Hiratsuka J., Shibata M.,  
 RA Wakasugi T., Sugitara M.;  
 RT "Chloroplast DNA of black pine retains a residual inverted repeat  
 RT lacking rRNA genes: nucleotide sequences of trnD, trnK, psbA, trnI and  
 RT trnH and the absence of rps16";

RL Mol. Gen. Genet. 232:206-214(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95094312; PubMed=8001170;  
 RA Tsudzuki J., Ito S., Tsudzuki T., Wakasugi T., Sugitara M.;  
 RT "A new gene encoding tRNA pro (GGG) is present in the chloroplast  
 RT genome of black pine: a compilation of 32 tRNA genes from black pine  
 RT chloroplasts";  
 RL Curr. Genet. 26:153-158(1994).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95024047; PubMed=7937893;  
 RA Wakasugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T.,  
 RA Sugitara M.;  
 RT "Loss of all ndh genes as determined by sequencing the entire  
 RT chloroplast genome of the black pine Pinus thunbergii";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798(1994).  
 DR EMBL; D17510; BAA04433.1; -  
 KW Chloroplast.  
 SQ SEQUENCE 42 AA; 4480 MW; 0B5B2669DB4AD8A4 CRC64;

Query Match 11.3%; Score 50; DB 8; Length 42;  
 Best Local Similarity 37.5%; Pred. No. 42;  
 Matches 6; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 63 CKLEPEPRIMVVPAL 78  
 DB 6 CSSAPPEMWTIOGTL 21

RESULT 11

OY 090602 PRELIMINARY; PRT; 59 AA.  
 AC 090602;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE Grapevine leafroll-associated virus 1 helicase (HEL) and RNA-dependent  
 DE RNA polymerase (POL) genes; and p7, HSP70-like protein, p55, coat  
 DE protein (CP), p55 (CPd1), p50 (CPd2), p22, and p24 genes.  
 OS grapevine leafroll-associated virus 1.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;  
 OC Closterovirus.  
 OX NCBI\_TaxID=47965;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20141373; PubMed=10675398;  
 RA Fazell C.F., Rezaian M.A.;  
 RT "Nucleotide sequence and organization of ten open reading frames in  
 RT the genome of Grapevine leafroll-associated virus 1 and identification  
 RT of three subgenomic RNAs";  
 RL J. Gen. Virol. 81:605-615(2000).  
 DR EMBL; AF195822; AAF22739.1; -  
 KW SEQUENCE 59 AA; 6736 MW; B90BED563F81DCF2 CRC64;

Query Match 11.3%; Score 50; DB 12; Length 59;  
 Best Local Similarity 25.9%; Pred. No. 59;  
 Matches 15; Conservative 10; Mismatches 19; Indels 14; Gaps 2;

OY 3 LVLVSLCLLILCFSTFEGKRRPAKAMSGRTRLCCHRV--PSPNSTNLKGHV 57  
 DB 10 LVTVSLFVILVLCFAY-----FVRAIRHCCAKIEDKPVGGFTNSFRIV 56

RESULT 12

OY 091KHI PRELIMINARY; PRT; 76 AA.  
 AC 091KHI;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Polyprotein (Fragment).  
 OS Hepatitis C virus.

```

OC  Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC  Hepatitisvirus
OC  NCBI_TaxID=11103;
OX  (1)
RN  SEQUENCE FROM N.A.
RC  STRAIN=133;
RA  Gimenez-Barcos M., Franco S., Suarez Y., Fornis X., Ampurdanes S.,
RA  Puig-Basagott F., Sanchez-Fueyo A., Barrera J.M., Lloret J.M.,
RA  Bruix J., Sanchez-Tapias J.M., Rodas J., Saliz J.C.,
RT  "High amino acid variability within the NS5A gene of hepatitis C virus
RT  (HCV) is associated with hepatocellular carcinoma in patients with
RT  HCV-1b related cirrhosis.";
RL  Hepatology 0:0-0(2001).
DR  EMBL: AF379099; AAK63318.1; -
DR  InterPro: IPR002522; HCV_capsid.
DR  Pfam: PF01543; HCV_capsid; 1.
FT  NON_TER
FT  1
FT  76
SQ  SEQUENCE 76 AA; 8660 MW; 394FC42024A5B4A8 CRC64;

Query Match 11.3%; Score 50; DB 12; Length 76;
Best Local Similarity 33.3%; Pred. No. 75;
Matches 14; Conservative 4; Mismatches 16; Indels 8; Gaps 2;

OY  35 RRRLCHRVSPNSNTLKGHHVRLCKPKLEPEPLMYVPG 76
DB  36 RARRKTSER-----SNPRGRPPISKAC--QPERTWHPG 69

RESULT 13
O9JHY4 PRELIMINARY; PRT; 80 AA.
AC  O9JHY4;
DT  01-OCT-2000 (TREMBLrel. 15, Created)
DT  01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT  01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE  Elatin-like protein I (SWAM1 protein).
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
OX  NCBI_TaxID=10090;
RN  (1)
RP  SEQUENCE FROM N.A.
RC  STRAIN=129/SVJ;
RA  Hagwaera K., Kikuchi T., Endo Y., Hugin, Takahashi M., Xin X.,
RA  Hoshi S., Miki M., Inooka N., Tokue Y., Nuklwa T.,
RT  "Mouse SWAM1 (single WAP motif protein 1) gene.";
RT  Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
RL  EMBL: AF276974; AAF86471.1; -
DR  EMBL: AF483009; AAL90747.1; -
DR  HSSP: 046655; ICDH.
DR  InterPro: IPR002221; WAP.
DR  Pfam: PF00095; wap; 1.
DR  PRINTS: PR00003; 4DISULPHCORE.
DR  SMART: SM00217; WAP; 1.
DR  PROSITE: PS00317; 4.DISULFIDE_CORE; 1.
SQ  SEQUENCE 80 AA; 9237 MW; 9135647ED91F63ED CRC64;

Query Match 11.2%; Score 49.5; DB 11; Length 80;
Best Local Similarity 34.7%; Pred. No. 92;
Matches 17; Conservative 8; Mismatches 17; Indels 7; Gaps 2;

OY  1 MRLVLSLLCLLLCFSTFGKRRPAKAWGRRTRLC--HRRPSP 47
DB  1 MKLLGSLAVTLCCNNARPEIKKNVFSKPG-----YCPERYVPCP 44

RESULT 14

```

```

O9YU66 PRELIMINARY; PRT; 71 AA.
ID  O9YU66;
AC  O9YU66;
DT  01-MAY-2000 (TREMBLrel. 13, Created)
DT  01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT  01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE  CG10714 protein.
GN  LY OR CG10714.
OS  Drosophila melanogaster (Fruit fly).
OC  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC  Ephydroidea; Drosophilidae; Drosophila.
OX  NCBI_TaxID=7227;
RN  (1)
RP  SEQUENCE FROM N.A.
RC  STRAIN=BERKELEY;
RC  MEDLINE=20196006; PubMed=10731132;
RA  Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA  Adams M.D., Celniker S.E., Holt R.A., Hoskins R.A., Galle R.F.,
RA  Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Henderson S.N.,
RA  George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA  Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA  Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA  Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mills G.L.G.,
RA  Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA  Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA  Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA  Borokova D., Botchan M.R., Bouck J., Brockschtein P., Brotler P.,
RA  Burks K.C., Busam D.A., Butler H., Cadieu E., Center H., Chandra I.,
RA  Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA  de Pablos B., Delcher A., Deng Z., Dey A.D., Dew I., Dietz S.M.,
RA  Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA  Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA  Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA  Glosik A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA  Haritz N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA  Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwam C.,
RA  Jalali M., Kalush F., Karpen G.H., Ke Z., Kenton J.A., Ketchum K.A.,
RA  Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA  Laszo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA  Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA  Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA  Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA  Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleib J.M.,
RA  Palazolo M., Pittman G.S., Pan S., Pollard J., Pui V., Reese M.G.,
RA  Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA  Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA  Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA  Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA  Wang Z.-Y., Wattarman D.A., Weinstein G.M., Weissbach J.,
RA  Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA  Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao X., Zheng L.,
RA  Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Zhang H.O.,
RA  Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT  "The genome sequence of Drosophila melanogaster.";
RL  Science 287:2185-2195(2000).
DR  EMBL: AF003538; -; NOT_ANNOTATED_CDS.
DR  FlyBase: FBgn0002573; Ly.
DR  FLYBASE: FBgn0002573; Ly.
SQ  SEQUENCE 71 AA; 7887 MW; D5AEF002A90AFC47 CRC64;

Query Match 11.0%; Score 48.5; DB 5; Length 71;
Best Local Similarity 29.6%; Pred. No. 11e+02;
Matches 16; Conservative 0; Mismatches 15; Indels 23; Gaps 2;

OY  30 KAWSRR-----TRLCCHRVSPNSNTLKGHHVRLCKPKLEPEPLMYV 73
DB  29 KQWQRAASAFNTICPFCTRCSTRVP-----YCGPFCPCPRRLMY 69

RESULT 15
O61910 PRELIMINARY; PRT; 61 AA.
ID  O61910;
AC  O61910;
DT  01-NOV-1996 (TREMBLrel. 01, Created)

```

DT 01-NOV-1996 (T-EMBLrel. 01, last sequence update)  
 DT 01-DEC-2001 (T-EMBLrel. 19, last annotation update)  
 DE Mammary transforming protein.  
 GN PEAL5 OR MAT1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BALB/C; TISSUE-BREAST TUMOR;  
 RX MEDLINE-95024046; PubMed-7937892;  
 RA Bera T.K., Guzman R.C., Miyamoto S., Panda D.K., Sasaki M., Hanyu K.,  
 RA Enami J., Nandi S.;  
 RT "Identification of a mammary transforming gene (MAT1) associated with  
 RT mouse mammary carcinogenesis."  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:9789-9793(1994).  
 DR EMBL; L31958; AAC37665.1; -  
 DR MGD; MGI:104799; Pea15.  
 SQ SEQUENCE 61 AA; 7137 MW; DDC1F91842892171 CRC64;  
 Query Match 10.9%; Score 48; DB 11; Length 61;  
 Best Local Similarity 28.3%; Pred. No. 1,1e+02;  
 Matches 13; Conservative 7; Mismatches 16; Indels 10; Gaps 2;  
 QY 17 FSIFSTEGKRRPAKANGR-----RTRLCCHRVSPSPNSTNLKGHH 56  
 DB 13 FVVFISINLSRPEREWGMPQKSGRAKL---LQSPNRKHISTNH 54

Search completed: May 8, 2003, 21:47:37  
 Job time : 33 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 21:42:53 ; Search time 10 Seconds

(without alignments)  
335,958 Million cell updates/sec

Title: US-09-599-087B-5

Perfect score: 442  
Sequence: 1 MRLVLSLLCLLCLFCSIF.....PCKLEPERLWVPGALQV 81

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 7253

Minimum DB seq length: 25  
Maximum DB seq length: 80

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59	13.3	55	1 ATP8_LATCH	003168 latimeria c
2	53	12.0	56	1 RS14_PYRHO	074093 pyrococcus
3	48.5	11.0	78	1 R19_PROWI	P46750 protheca
4	48	10.9	61	1 R14A_LACLA	Q9c4t5 iactococcus
5	48	10.9	74	1 SG57_DROME	P02845 drosophila
6	47	10.6	64	1 BD01_SHEEP	O19038 ovis aries
7	47	10.6	69	1 RS7_MERMT	O93636 methanococ
8	46.5	10.5	69	1 BD01_MOUSE	P56386 mus musculu
9	46	10.4	69	1 BD01_RAT	O89117 rattus norv
10	45	10.2	75	1 SG58_DROME	P02842 drosophila
11	44.5	10.1	46	1 AX2_BETVU	P82010 beta vulgar
12	44.5	10.1	61	1 R14A_STRPY	Q9a1w1 streptococ
13	44.5	10.1	75	1 VGE_BPPHK	Q38040 bacterioph
14	44	10.0	50	1 HSP1_GORGO	P35303 gorilla gor
15	44	10.0	59	1 SCK2_ANDMA	P45696 androcton
16	44	10.0	63	1 ITR4_LUCFY	P34950 iuffa cylin
17	43.5	9.8	75	1 VGE_BPAL3	P31280 bacterioph
18	43	9.7	49	1 HSP1_RABIT	P10119 oryctolagu
19	43	9.7	63	1 FCAL_CRIFA	Q23688 crithidia f
20	43	9.7	78	1 L2MU_ADECC	O65993 canine aden
21	43	9.7	75	1 YCEK_ECOLI	P45806 escherichia
22	42.5	9.6	54	1 RL37_TREVO	O97b6 thermoplas
23	42.5	9.6	55	1 NXB4_CERLA	P01535 ceratratulu
24	42.5	9.6	57	1 RL32_STRAM	O99ux6 staphylococ
25	42.5	9.6	66	1 CXAX_CONTE	O9xk6 conus texti
26	42.5	9.6	69	1 CXAX_CONTE	O9xk6 conus texti
27	42.5	9.6	72	1 VG18_BPMU	O38625 bacterioph
28	42.5	9.6	80	1 AFP2_RAPSA	P30230 raphanus sa
29	42	9.5	49	1 HSP1_SAGIM	P24714 saginatus im
30	42	9.5	50	1 HSP1_BOVIN	P02318 bos taurus
31	42	9.5	50	1 HSP1_SHEEP	P04102 ovis aries
32	42	9.5	50	1 PND1_ECOLI	P11902 escherichia
33	42	9.5	70	1 YVAK_VACCC	P20520 vaccinia vl

34	41.5	9.4	53	1 RS14_METVA	P14041 methanococ
35	41.5	9.4	64	1 LAP_BOVIN	Q28680 bos taurus
36	41	9.3	27	1 CXM3_CONPU	P58925 conus purpu
37	41	9.3	52	1 V07K_FYMY	P22171 foxtail mos
38	41	9.3	63	1 RL35_CAMJE	O9p1q1 campylobact
39	41	9.3	70	1 RS21_RALSO	O8y3h2 ralsionia s
40	41	9.3	72	1 NX11_DENJA	P61393 dendroaspis
41	41	9.3	79	1 FRD1_AOUAE	O67251 aquilex aeo
42	40.5	9.2	54	1 RL37_TREAC	O9h1t7 thermoplas
43	40.5	9.2	61	1 R21_BP933	O9xj17 bacterioph
44	40.5	9.2	77	1 LEA2_MACMU	O95m25 macaca mula
45	40.5	9.2	77	1 Y125_HUMAN	Q14138 homo sapien

## ALIGNMENTS

RESULT 1	ATP8_LATCH	STANDARD;	PRT;	55 AA.
ID	ATP8_LATCH			
AC	003168;			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	ATP synthase protein 8 (EC 3.6.3.14) (ATPase subunit 8) (A6L).			
GN	MTATP8 OR ATP8.			
OS	Latimeria chalumnae (Latimeria) (Coelacanth).			
OC	Mitochondrion.			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Coelacanthiformes; Coelacanthidae; Latimeria.			
OX	NCBI_TaxID=7897;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Zardoya R., Meyer A.;			
RL	Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.			
CC	- FUNCTION: THIS IS ONE OF THE CHAINS OF THE NONENZYMATIC COMPONENT			
CC	(CF1O) SUBUNIT OF THE MITOCHONDRIAL ATPASE COMPLEX.			
CC	- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate +			
CC	H(+) (out).			
CC	- SUBCELLULAR LOCATION: Membrane-bound.			
CC	- SIMILARITY: BELONGS TO THE ATPASE PROTEIN 8 FAMILY.			
CC	- This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
DR	EMBL: U82228; AAC60322.1;			
DR	InterPro: IPR001421; ATPase8_mt.			
KW	Pfam: PF00895; ATP-synt_8; 1.			
DR	Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.			
FT	TRANSMEM 4 24			
FT	POTENTIAL.			
SO	SEQUENCE 55 AA; 6569 MW; 7FE36319EAF625B CRC64;			
Query Match	13.3%;	Score 59;	DB 1;	Length 55;
Best Local Similarity	34.3%;	Pred. No. 3.9;		
Matches 23;	Conservative 7;	Mismatches 15;	Indels 22;	Gaps 5;
QY	7 SLLCLILLCLFSTGKRRPAKAWGRRPRLCCHVPS-PNSTNLKGVHRLCKPCKL 65			
DB	7 SPWLLILLESWLLFLT--MLPSK-----TOL-HTFPMPSIQNM-----CKQ 45			
QY	66 EPEPRLW 72			
DB	46 EPEPWTW 52			
RESULT 2	RS14_PYRHO	STANDARD;	PRT;	56 AA.
ID	RS14_PYRHO			

AC 074093;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 30S ribosomal protein S14P.  
 GN RPS14P OR PHS047 OR PAB7080.  
 OS Pyrococcus horikoshii, and  
 OS Pyrococcus abyssi.  
 OC Archaea: Euryarchaeota: Thermococci: Thermococcales: Thermococcaceae;  
 OC Pyrococcus.  
 NCBI\_TaxID=53953, 29292;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-P. horikoshii; STRAIN-OT3;  
 RX MEDLINE=98344137; PubMed=9679194;  
 RA Kawarabayashi Y., Sawada M., Horikawa H., Halkawa Y., Hino Y.,  
 RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
 RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Yamazaki J., Kusuda N., Oguchi A.,  
 RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kusuda N., Oguchi A.,  
 RA Masuchi Y., Shizuya H., Kikuchi H.,  
 RT "Complete sequence and gene organization of the genome of a hyper-  
 RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
 RL DNA Res. 5:55-76(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-P. abyssi; STRAIN-GE5 / Orsay;  
 RA Hellig R.;  
 RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome  
 RT structure and evolution.";  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: BELONGS TO THE S14P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb-sib.ch](mailto:license@sdb-sib.ch)).  
 CC -----  
 DR EMBL: AP000007; BAA30879.1; -  
 DR EMBL: AJ248284; CAB49249.1; -  
 DR InterPro: IPR001209; Ribosomal\_S14.  
 DR Pfam: PF00253; Ribosomal\_S14; 1.  
 DR PROSITE: PS00527; RIBOSOMAL\_S14; 1.  
 KW Ribosomal protein; Complete proteome.  
 SQ SEQUENCE 56 AA; 6626 MW; B5D97E9E2841F686 CRC64;  
 Query Match 12.0%; Score 53; DB 1; Length 56;  
 Best Local Similarity 29.8%; Pred. No. 18;  
 Matches 14; Conservative 8; Mismatches 21; Indels 4; Gaps 1;  
 OY 25 KRBPAAWMSGRRLCCHRPSPNSNLKGHVRLCKPCKLEPERL 71  
 DB 8 KRBPKEFGKGRRCRCRGOT---GPIIRLHGLMLCRHCRREVAPRL 50  
 RESULT 3  
 RT19\_PROTMI STANDARD: PRT; 78 AA.  
 AC P46750;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Mitochondrial ribosomal protein S19.  
 GN RPS19.  
 OS Prototheca wickerhamii.  
 OC Mitochondrion.  
 OC Eukaryota: Viridiplantae: Chlorophyta: Trebouxiophyceae; Chlorellales;  
 OC Chlorellaceae; Prototheca.  
 NCBI\_TaxID=3111;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-263-11;  
 RX MEDLINE=94180393; PubMed=8133522;  
 RA Wolff G., Plante I., Lang B.F., Knecht U., Burger G.,  
 RT "Complete sequence of the mitochondrial DNA of the chlorophyte alga  
 RT Prototheca wickerhamii. Gene content and genome organization.";  
 RL J. Mol. Biol. 237:75-86(1994).  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial.  
 CC -1- SIMILARITY: BELONGS TO THE S19P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb-sib.ch](mailto:license@sdb-sib.ch)).  
 CC -----  
 DR EMBL: U02970; AAD12637.1; -  
 DR HSSP: P80381; 10KF.  
 DR InterPro: IPR002222; Ribosomal\_S19.  
 DR Pfam: PF00203; Ribosomal\_S19; 1.  
 DR PRINTS: PR00975; RIBOSOMAL\_S19.  
 DR ProDom: PD001012; Ribosomal\_S19; 1.  
 DR TIGRfam: TIGR01050; rpsS\_bact; 1.  
 DR PROSITE: PS00523; RIBOSOMAL\_S19; 1.  
 KW Ribosomal protein; Mitochondrion.  
 SQ SEQUENCE 78 AA; 8847 MW; 9B53BDA9A287BCDB CRC64;  
 Query Match 11.0%; Score 48.5; DB 1; Length 78;  
 Best Local Similarity 33.3%; Pred. No. 77;  
 Matches 13; Conservative 7; Mismatches 14; Indels 5; Gaps 2;  
 OY 30 KASGRRRLCCHRPSPNSNLKGHVRLCKPCKLEPE 68  
 DB 23 KIMS-BRSAL---LQFVGKTVSHNGRIPIFKISPE 56  
 RESULT 4  
 R14A\_LACLA STANDARD: PRT; 61 AA.  
 AC Q9CDX5;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE 30S ribosomal protein S14-1.  
 GN RPSN1 OR IL2086.  
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
 OC Bacteria: Firmicutes: Lactobacillales; Streptococcaceae; Lactococcus.  
 NCBI\_TaxID=1360;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-IL1403;  
 RX MEDLINE=11235186; PubMed=11337471;  
 RA Bolotin A., Mincker P., Manger S., Jallion O., Malame K.,  
 RA Weissendbach J., Enrich S.D., Sorokin A.,  
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus  
 RT lactis ssp. lactis IL1403.";  
 RL Genome Res. 11:731-753(2001).  
 CC -1- FUNCTION: Known to be required for the assembly of 30S particles  
 CC and may also be responsible for determining the conformation of  
 CC the 16S rRNA at the A site (by similarity).  
 CC -1- SIMILARITY: BELONGS TO THE S14P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb-sib.ch](mailto:license@sdb-sib.ch)).  
 CC -----  
 DR EMBL: AE006437; AAK06184.1; -

RA Ye J., Ieh R.-F., Zaveli J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Ra X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gblbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
CC -1- DEVELOPMENTAL STAGE: PRODUCED BY THIRD-INSTAR LARVAE.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial/  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X01918; CAA25993.1; .  
DR EMBL; AE003544; AAF60058.1; .  
DR PIR; A03330; GSFF7.  
DR FlyBase; FBgn0003377; Sgs7.  
KW Signal.  
FT SIGNAL  
FT CHAIN  
FT FT 1 23 SALIVARY GLUE PROTEIN SGS-7.  
SQ SEQUENCE 74 AA; 7919 MW; 914ACASDB9153E29 CRC64;  
  
Query Match 10.9%; Score 48; DB 1; Length 74;  
Best Local Similarity 26.8%; Pred. No. 84;  
Matches 22; Conservative % 18; Mismatches 26; Indels 16; Gaps 7;  
  
QY 1 MRLLVLSLLCILLICFEIFSTEG--KRRP---AKAMSG--RRTRLCCHRPSPNSNLT 52  
Db |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
1 MKLIAVTIACILILIGFDLDLGAGCECPGCGPGGACTGCPEKPQLCOLISDIRLQQ 60  
  
QY 53 KGHVRYLCKPKCLKEEPRLMVV 74  
Db |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|:  
61 K---IRKC-VCG-EPO---WMI 74  
  
RESULT 6  
BD01\_SHEEP STANDARD; PRF; 64 AA.  
AC 019038;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Beta-defensin 1 precursor (BD-1) (SBD1).  
GN DEFBL.  
OS Ovis aries (Sheep).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Euthera; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Caprinae; Ovis.  
OX NCBI\_TaxID=9940;  
[1]  
SEQUENCE FROM N.A.  
RX MEDLINE=98138497; PubMed=9478010;  
RA Hutner K.M., Brezinski-Caliguri D.J., Mahoney M.W., Diamond G.;  
RT "Anticicrobial peptide expression is developmentally regulated in the  
RT ovine gastrointestinal tract.";  
RL J. Nutr. 128:2975-2995(1998).  
[2]  
SEQUENCE FROM N.A.  
RP TISSUE=Trochaea;  
RC MEDLINE=98121317; PubMed=9461419;  
RX Hutner K.M., Lambeth M.R., Burkin H.R., Broad T.E.;  
RT "Localization and genomic organization of sheep antimicrobial peptides  
RT genes.";  
RL Gene 206:85-91(1998).  
CC -1- FUNCTION: HAS BACTERICIDAL ACTIVITY (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).  
CC -1- SIMILARITY: BELONGS TO THE BETA-DEFENSIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: U75250; AAB61995.1; -  
DR HSSP; P46170; IBB.  
DR InterPro: IPR001855; Defensin\_beta.  
DR InterPro: IPR001271; Defensin\_mammal.  
DR Pfam: PF00711; Defensin\_beta; 1.  
DR SMART; SM00048; DEFSN; 1.  
KW Antibiotic; Signal.  
FT SIGNAL 1 22 BY SIMILARITY.  
FT CHAIN 23 64 BETA-DEFENSIN 1.  
FT DISULFID 31 60 BY SIMILARITY.  
FT DISULFID 38 53 BY SIMILARITY.  
FT DISULFID 43 61 BY SIMILARITY.  
SQ SEQUENCE 64 AA; 7244 MW; 3529A9B76ABD023A CRC64;

Query Match 10.6%; Score 47; DB 1; Length 64;  
Best Local Similarity 30.2%; Pred. No. 95;  
Matches 19; Conservative 7; Mismatches 19; Indels 18; Gaps 4;

QY 12 ILLCFSIFSTEGKRRPAKANGRTLCCHR-----VPS---PNSYLNKGHVRACRPRCK 64  
DB 6 LLVLFVFLVLAG-----SGFTQGVNRNLSCHRNKGVCVPSRCP-----RHMQIGTCR 54  
QY 65 LEP 67  
DB 55 GPP 57

RESULT 7  
RS7\_METMT STANDARD; PRT; 69 AA.  
AC 093636;  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE 30S ribosomal protein S7P (Fragment).  
GN RPS7P OR S7.  
OS Methanococcoides methylans.  
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;  
CC Methanosarcinaceae; Methanococcoides.  
OX NCBI\_Taxid=2226;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DSM 26577;  
RX MEDLINE=99059471; PubMed=9845338;  
RA Thomas T., Cavicchioli R.;  
RT "Archaeal cold-adapted proteins: structural and evolutionary analysis  
RT of the elongation factor 2 proteins from psychrophilic, mesophilic and  
RT thermophilic methanogens.";  
RL FEBS Lett. 439:281-287(1998).  
CC -1- FUNCTION: PROTEIN S7 BINDS SPECIFICALLY TO PART OF THE 3' END OF  
CC 16S RIBOSOMAL RNA (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE S7P FAMILY OF RIBOSOMAL PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AF022779; AAC79196.1; -  
DR InterPro: IPR000235; Ribosomal\_S7.  
DR Pfam: PF00177; Ribosomal\_S7; 1.  
DR PROSITE: PS00052; RIBOSOMAL\_S7; PARTIAL.  
KW RIBOSOMAL protein; rRNA-binding.  
FT NON\_TER 1 1  
SQ SEQUENCE 69 AA; 7488 MW; 7E64F12BDA024160 CRC64;

Query Match 10.6%; Score 47; DB 1; Length 69;  
Best Local Similarity 62.5%; Pred. No. 1e+02;  
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 16 CFSIFSTEGKRRPAKA 31  
DB 52 CFSINRRDKERPAKA 67

RESULT 8  
BD01\_MOUSE STANDARD; PRT; 69 AA.  
ID BD01\_MOUSE  
AC P56386;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Beta-defensin 1 precursor (Bd-1) (mbd-1).  
GN DEFB1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Merritt M., Hillier L., Allen M., Bowles M., Dietrich N., Dubuque T.,  
RA Geisel S., Kucaba T., Lacy M., Le M., Martin J., Morris M.,  
RA Schellenberg K., Stepien M., Tan F., Underwood K., Moore B.,  
RA Theising B., Wylie T., Lennon G., Soares B., Wilson R., Waterston R.;  
RA Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97431609; PubMed=9287114;  
RA Hultner K.M., Kozak C.A., Bevins C.L.;  
RT "The mouse genome encodes a single homolog of the antimicrobial  
RT peptide human beta-defensin 1.";  
RL FEBS Lett. 413:45-49(1997).  
RN [3]  
RP TISSUE SPECIFICITY.  
RC STRAIN=C57BL/6, 129/SVJ, and FVB; TISSUE=Lung;  
RX MEDLINE=20517883; PubMed=10922379;  
RA Jia H.P., Nwok S.A., Schutte B.C., Lee S.K., Vivado A., Tack B.F.,  
RA Bevins C.L., McCray P.B. Jr.;  
RT "A novel murine beta-defensin expressed in tongue, esophagus, and  
RT trachea.";  
RL J. Biol. Chem. 275:33314-33320(2000).  
CC -1- FUNCTION: HAS BACTERICIDAL ACTIVITY (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).  
CC -1- TISSUE SPECIFICITY: DETECTED IN KIDNEY.  
CC -1- SIMILARITY: BELONGS TO THE BETA-DEFENSIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AA071157; -; NOT\_ANNOTATED\_CDS.  
DR EMBL: AA065510; -; NOT\_ANNOTATED\_CDS.  
DR EMBL: AA108061; -; NOT\_ANNOTATED\_CDS.  
DR EMBL: AA107538; -; NOT\_ANNOTATED\_CDS.  
DR EMBL: AA105324; -; NOT\_ANNOTATED\_CDS.  
DR EMBL: AF003525; AAB72003.1; JOINED.  
DR EMBL: AF003524; AAB72003.1; JOINED.  
DR MGI: 1096878; Defb1.  
DR InterPro: IPR001855; Defensin\_beta.  
DR InterPro: IPR001271; Defensin\_mammal.  
DR Pfam: PF00711; Defensin\_beta; 1.  
DR SMART; SM00048; DEFSN; 1.  
KW Antibiotic; Signal.  
FT SIGNAL 1 21 POTENTIAL.



```

FT PROPEP      22      32      BY SIMILARITY.
FT CHAIN       33      69      BETA-DEFENSIN 1.
FT DISULFID    37      66      BY SIMILARITY.
FT DISULFID    44      59      BY SIMILARITY.
FT DISULFID    49      67      BY SIMILARITY.
SQ SEQUENCE    69 AA; 7749 MW; 8B12BD3BFB3ACE4 CRC64;

Query Match
Best Local Similarity 31.1%; DB 1; Length 69;
Matches 19; Conservative 8; Mismatches 21; Indels 13; Gaps 4;

OY 12 ILICFSITSTGKRRPRAPAMSGRT--RLC-----CHRVSPNSTNKGHHVRLCKPC 63
DB 8 LVWICF-LFSQMEPGVIGLISLRRTDYKCLQHGSGFCLRSSCPSTKLG-----TCKPD 62

OY 64 K 64
DB 63 K 63

RESULT 9
BD01_RAT
ID BD01_RAT STANDARD; PRT; 69 AA.
AC 089117;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Beta-defensin 1 precursor (BD-1) (RBD-1).
GN DEFBL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Goto-Kakizaki; TISSUE-Kidney;
RA Page R.A., Malik A.N.;
RT "Rat beta defensin-1 peptide: candidate marker for diabetic
  neuropathy."
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Mistral;
RC MEDLINE-99386883; PubMed-10456937;
RA Jia H.P., Mills J.N., Barahmand-Pour F., Nishimura D.,
RA Mallampalli R.K., Wang G., Miles K., Tack B.F., Bevins C.L.,
RA McCreay P.B., Jr.;
RT "Molecular cloning and characterization of rat genes encoding
  homologues of human beta-defensins."
RL Infect. Immun. 67:4827-4833(1999).
CC -1- FUNCTION: HAS BACTERICIDAL ACTIVITY (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: Highly expressed in kidney.
CC -1- SIMILARITY: BELONGS TO THE BETA-DEFENSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation-
  CC the European Bioinformatics Institute. There are no restrictions on its
  CC use by non-profit institutions as long as its content is in no way
  CC modified and this statement is not removed. Usage by and for commercial
  CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
  CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF093536; AAC61871.1; -
DR EMBL: AF068860; AAC28071.1; -
DR InterPro: IPR001855; Defensin_beta.
DR InterPro: IPR001271; Defensin_mammal.
DR Pfam: PF00711; Defensin_beta; 1.
DR SMART: SM00048; DEFSN; 1.
KW Antibiotic; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT PROPEP 22 32 BY SIMILARITY.
FT CHAIN 33 69 BETA-DEFENSIN 1.

```

```

FT DISULFID    37      66      BY SIMILARITY.
FT DISULFID    44      59      BY SIMILARITY.
FT DISULFID    49      67      BY SIMILARITY.
SQ SEQUENCE    69 AA; 7837 MW; 66B1F0C29B5BC991 CRC64;

Query Match
Best Local Similarity 31.1%; DB 1; Length 69;
Matches 19; Conservative 6; Mismatches 24; Indels 12; Gaps 3;

OY 12 ILICFSITSTGKRRPRAPAMSGRT--RLC-----CHRVSPNSTNKGHHVRLCKPC 63
DB 7 LVWLFELFSPQMELAGLISLRRTDYKCLQHGSGFCLRSSCPSTKLG-----TCKPD 62

OY 64 K 64
DB 63 K 63

RESULT 10
SGS8_DROME
ID SGS8_DROME STANDARD; PRT; 75 AA.
AC P02842; Q9VT19;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Salivary glue protein Sgs-8 precursor.
GN SGS8 OR C66132.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE-83294545; PubMed-6411930;
RA Garfinkel M.D., Pultz R.E., Meyerowitz E.M.;
RT "DNA sequences, gene regulation and modular protein evolution in the
  RT Drosophila 68C glue gene cluster."
RL J. Mol. Biol. 168:765-789(1983).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Berkeley;
RC MEDLINE-20196006; PubMed-10731132;
RA Adams M.D., Gelunker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Adavani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktiroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borikova D., Bousam D.A., Bouck J., Broxstein P., Brotler P.,
RA Burks K.C., Busam D.R., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Flatschman W.,
RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Giodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibergam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mewklov G., Mlshina N.V., Moadary C., Morris J., Moshirefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Pui Y., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

```

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin C.M., Venter J.C.,  
 RT "The genome sequence of *Drosophila melanogaster*,"  
 RL Science 287:2185-2195(2000).  
 CC -1- DEVELOPMENTAL STAGE: PRODUCED BY THIRD-INSTAR LARVAE.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: X01918; CA25992.1; -  
 DR EMBL: AE003544; AAF50059.1; -  
 DR PIR: A03331; GSF8.  
 DR Flybase: FBgn0003378; Sgs8.  
 KM Signal.  
 FT STGNL 1 24  
 FT CHAIN 25 75 SALIVARY GLUE PROTEIN SGS-8.  
 SQ SEQUENCE 75 AA; 7917 MW; 27B818DFE138C636 CRC64;  
 Query Match 10.2%; Score 45; DB 1; Length 75;  
 Best Local Similarity 44.4%; Pred. No. 1.8e+02;  
 Matches 8; Conservative 7; Mismatches 3; Indels 0; Gaps 0;  
 Oy 1 MRLVSSLLCILLCS 18  
 Db 1 MKLVAVIACIMIGFA 18  
 ID AX2.BETVU STANDARD; PRT; 46 AA.  
 AC P82010; P81510;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Antifungal protein AX2.  
 OS Beta vulgaris (Sugar beet).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Caryophyllales; Caryophyllaceae; Beta.  
 NCBI\_TaxID=3555.  
 RN [1]  
 RP SEQUENCE.  
 RC STRAIN=cv. Turbo, and cv. Rhizor; TISSUE=leaf;  
 RA MEDLINE=95383713; Pubmed=7655063;  
 RA Krach K.M., Nielsen J.E., Nielsen K.K., Dreboldt S., Mikkelsen J.D.,  
 RT "Characterization and localization of new antifungal cysteine-rich  
 RT proteins from *Beta vulgaris*," 8:424-434(1995).  
 RL Mol. Plant Microbe Interact. 8:424-434(1995).  
 CC -1- FUNCTION: STRONG INHIBITING ACTIVITY AGAINST *C. BETICOLOA* AND OTHER  
 CC FILAMENTOUS FUNGI. LITTLE OR NO EFFECT AGAINST BACTERIA.  
 CC -1- TISSUE SPECIFICITY: LEAVES AND FLOWERS.  
 CC -1- MASS SPECTROMETRY: MW=5193; MW ERR=3; METHOD=Electrospray.  
 CC -1- SIMILARITY: BELONGS TO THE PLANT DEFENSIN FAMILY.  
 DR HSP; P20158; IGPS.  
 DR Interpro: IPR002118; Gamma-thionin.  
 DR Interpro: IPR003614; Knote1.  
 DR Pfam: PF00304; Gamma-thionin; 1.  
 DR Prodom: PD002594; Gamma-thionin; 1.  
 DR SMART: SM00505; Knote1.1.  
 KM Plant defense; Antibiotic; Fungicide.  
 FT DISULFID 3 46  
 FT DISULFID 14 34 BY SIMILARITY.  
 FT DISULFID 20 40 BY SIMILARITY.  
 FT DISULFID 24 42 BY SIMILARITY.

SQ SEQUENCE 46 AA; 5185 MW; 9A536FE9E74B18A6 CRC64;  
 Query Match 10.1%; Score 44.5; DB 1; Length 46;  
 Best Local Similarity 27.9%; Pred. No. 1.3e+02;  
 Matches 12; Conservative 6; Mismatches 20; Indels 5; Gaps 1;  
 Oy 26 RRPAAKMSG-----RRRLCCHVSPSPNSYLNKGHHVRLCKPC 63  
 Db 4 RKPMSYFSGACFSDTNCOKACNREDWPNKCLGFKCECORPC 46  
 ID R14A.STRPY STANDARD; PRT; 61 AA.  
 AC Q9A1W1;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE 30S ribosomal protein S14-1.  
 GN RPSN1 OR SPY0064 OR SPYM18\_0065.  
 OS Streptococcus pyogenes (serotype M18).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 NCBI\_TaxID=1314, 186103;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SF370 / ATCC 700294 / Serotype M1;  
 RX MEDLINE=21192684; Pubmed=11296296;  
 RA Ferretti J.J., Mcshan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,  
 RA Sylva G.L., Sturdevant D.E., Rickielfs S.M., Porcella S.F.,  
 RA Primeaux C., Szate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,  
 RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,  
 RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.,  
 RA "Complete genome sequence of an M1 strain of *Streptococcus pyogenes*,"  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MGAS8232 / Serotype M18;  
 RX MEDLINE=21927593; Pubmed=11917108;  
 RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,  
 RA Sylva G.L., Sturdevant D.E., Rickielfs S.M., Porcella S.F.,  
 RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang O.,  
 RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.,  
 RT "Genome sequence and comparative microarray analysis of serotype M18  
 RT group A *Streptococcus* strains associated with acute rheumatic fever  
 RT outbreaks,"  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).  
 CC -1- FUNCTION: Known to be required for the assembly of 30S particles  
 CC and may also be responsible for determining the conformation of  
 CC the 16S rRNA at the A site (By similarity).  
 CC -1- SIMILARITY: BELONGS TO THE S14P FAMILY OF RIBOSOMAL PROTEINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: AE006478; AAK33195.1; -  
 DR EMBL: AE009959; AAL96889.1; -  
 DR Interpro: IPR001209; Ribosomal S14.  
 DR Pfam: PF00253; Ribosomal\_S14; 1.  
 DR ProSITE: PS00527; RIBOSOMAL\_S14; 1.  
 KM Ribosomal protein; Complete proteome.  
 SQ SEQUENCE 61 AA; 7073 MW; 3AE47B92DD4840BF CRC64;  
 Query Match 10.1%; Score 44.5; DB 1; Length 61;  
 Best Local Similarity 34.2%; Pred. No. 1.7e+02;  
 Matches 13; Conservative 6; Mismatches 14; Indels 5; Gaps 2;  
 Oy 26 RRPAAKMSGRRRLCCHVSPSPNSYLNKGHHVRLCKPC 63

```

Db      11 KRPAKHSIQAYTR--CEKCGPHSVYRK---FKLCRCV 43

RESULT 13
VGE_BPPHK STANDARD; PRT; 75 AA.
AC Q38040;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Lysis protein (E protein) (GPE).
GN E
OS Bacteriophage phi-K.
OC Viruses; ssDNA viruses; Microviridae; Microvirus.
OX NCBI_TaxID=10848;
RN [1]
RP SEQUENCE FROM N.A.
RA Kodaira K.-I., Tadokoro H., Suzuki K., Oki M., Manto S., Taketo A.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: E PROTEIN IS RESPONSIBLE FOR HOST CELL LYSIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X60323; CAA42889.1; -
KW Phage lysis protein.
SQ SEQUENCE 75 AA; 8441 MW; 7F900ECDC8A59045 CRC64;

Query Match 10.1%; Score 44.5; DB 1; Length 75;
Best Local Similarity 30.0%; Pred. No. 1,le+02;
Matches 21; Conservative 5; Mismatches 33; Indels 11; Gaps 3;

QY 3 LVLVSLILCLILCF--SIFSTEGKRPAKAMSGRRTRLCRRVSPNSTMLKGHVRLC 60
DB 13 LLLSLPLSLIMFIPSTF-----RQHSILMKARSLAKTLIMASARLPLSSRT--- 64
QY 61 KPCLLEPEPR 70
DB 65 -PCVLRQDSK 73

RESULT 14
HSP1_GORGO STANDARD; PRT; 50 AA.
AC P35303;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sperm protamine P1 (Cysteine-rich protamine).
GN PPM1.
OS Gorilla gorilla gorilla (Lowland gorilla).
OC Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Gorilla.
OX NCBI_TaxID=9595;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-9414943; Pubmed-8308910;
RA Relief J.D., Winkfein R.J., Dixon G.H., Adroer R., Queralt R.,
RA Ballabriga J., Oliva R.;
RT "Evolution of protamine P1 genes in primates.";
RL J. Mol. Evol. 37:426-434(1993).
CC -1- FUNCTION: PROTAGMINS SUBSTITUTE FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -1- SUBUNIT: CROSS-LINKED BY INTERCHAIN DISULFIDE BONDS AROUND THE
CC DNA-HELIX (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.

```

```

CC -1- TISSUE SPECIFICITY: TESTIS.
CC -1- SIMILARITY: BELONGS TO THE PROTAMINE P1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L14587; AAA51528.1; -
DR InterPro; IPR000221; Protamine_P1.
DR Pfam; PF00260; Protamine_P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
FT INTR. MET 0
FT INT. MET 0
SQ SEQUENCE 50 AA; 6600 MW; 0EE9ABB8451FEFA2 CRC64;

Query Match 10.0%; Score 44; DB 1; Length 50;
Best Local Similarity 32.1%; Pred. No. 1,le+02;
Matches 9; Conservative 3; Mismatches 16; Indels 0; Gaps 0;

QY 16 CFSIFSTEGKRPAKAMSGRRTRLCRR 43
DB 14 CYRQRTSRRRRRSRSCOTRRAMRCRR 41

RESULT 15
SCK2_ANDMA STANDARD; PRT; 59 AA.
AC P45696;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Kallitoxin 2 precursor (KTX 2).
GN KTX2.
OS Androctonus mauretanicus mauretanicus (Scorpion).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;
OC Butioidae; Butiidae; Androctonus.
OX NCBI_TaxID=6860;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-59.
RC TISSUE=Venom;
RX MEDLINE-95105165; Pubmed-7806508;
RA Laraba-Djebart F., Legros C., Crest M., Ceard B., Romi R.,
RA Mansuelle P., Jacquet G., van Rietsohoren J., Gola M., Rochat H.;
RT "The kallitoxin family enlarged. Purification, characterization, and
RT precursor nucleotide sequence of KTX2 from Androctonus australis
RT venom.";
RL J. Biol. Chem. 269:32835-32843(1994).
CC -1- FUNCTION: INHIBITOR OF THE DENDROTOXIN SENSITIVE VOLTAGE-DEPENDENT
CC K+ CHANNEL AND OF CALCIUM-ACTIVATED POTASSIUM CHANNELS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE SHORT SCORPION TOXIN FAMILY. POTASSIUM
CC CHANNEL INHIBITORS SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S74733; AAB33535.1; -
DR HSP; P55896; ISCO.
DR InterPro; IPR001947; Scorpion_toxins.
DR Pfam; PF00451; toxin_2; 1.
DR ProDom; PD003586; Scorpion_toxins; 1.
DR PROSITE; PS01138; SCORP_SHORT_TOXIN; 1.
KW Neurotoxin; Potassium channel inhibitor; signal.

```

```

FT SIGNAL 1 22
FT CHAIN 23 59 KALIOPOXIN 2.
FT DISULFID 29 49 BY SIMILARITY.
FT DISULFID 35 54 BY SIMILARITY.
FT DISULFID 39 56 BY SIMILARITY.
FT SITE 47 54 INTERACTION WITH CA(2+)-ACTIVATED K(+)
FT CHANNELS (POTENTIAL).
SQ SEQUENCE 59 AA; 6393 MW; DD3225C5D270EB26 CRC64;

Query Match 10.08; Score 44; DB 1; Length 59;
Best Local Similarity 23.3%; Pred. No. 1.9e+02;
Matches 14; Conservative 4; Mismatches 20; Indels 22; Gaps 1;

QY 5 VLSSLGLLILLCFSEFTEGKRRPAKAMSGRRRLCCHRVSPNSTNLKGHHVRLCKPCK 64
DB 3 VFSVLLILFLVCSMIGINAVRIPVSC-----KHSQCCKPCK 40

```

Search completed: May 8, 2003, 21:46:54  
 Job time : 12 secs





chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qiu, H.; Dragol, I.; Sellers, H.  
 1, R.R.; Melikjanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000  
 A>Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833; PMID:10952301  
 A:Accession: C82382

A:Status: preliminary  
 A:Molecule type: DNA

A:Residues: 1-79 <HEI>

A:Cross-references: GB:AE004432; GB:AE003853; NID:99658509; PIDN:AAF96960.1; GSPDB:GN001  
 A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VCA1066  
 A:Map position: 2

Query Match 11.5%; Score 51; DB 2; Length 79;  
 Best Local Similarity 42.9%; Pred. No. 1.4e+02;  
 Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 39 LCCHRVSPNSTNLKGHHVRL 59  
 | | | : | : | : | | | : |  
 Db 50 LTCHLPAPPSLSLGASHVKL 70

RESULT 8:  
 T07557

hypothetical protein 42f - Japanese black pine chloroplast  
 C:Species: chloroplast Pinus thunbergiana (Japanese black pine)

C:Date: 14-May-1999 #sequence\_revision 14-May-1999 #text\_change 18-Aug-2000  
 C:Accession: T07557

R:Nakatsuki, T.; Tsudzuki, J.; Ito, S.; Nakashima, K.; Tsudzuki, T.; Sugitara, M.  
 Proc. Natl. Acad. Sci. U.S.A. 91, 9794-9798, 1994

A>Title: Loss of all ndh genes as determined by sequencing the entire chloroplast genome  
 A:Reference number: Z16030; MUID:95024047; PMID:7937893

A:Accession: T07557

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-42 <NAK>

A:Cross-references: EMBL:DJ7510; NID:9529643; PIDN:BA04433.1; PID:q1262718  
 C:Genetics:

A:Genome: chloroplast  
 C:Keywords: chloroplast

Query Match 11.3%; Score 50; DB 2; Length 42;  
 Best Local Similarity 37.5%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 63 CKLEPRLMVVPGAL 78  
 | | | : | : | : | | | : |  
 Db 6 CSSAPDEKMTTIGTL 21

RESULT 9:  
 T11918

ribosomal protein S19 - Prototheca wickerhamii mitochondrion  
 C:Species: mitochondrion Prototheca wickerhamii

C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 26-Aug-1999  
 C:Accession: T11918

R:Wolff, G.; Plante, I.; Lang, B.F.; Kueck, U.; Burger, G.  
 J. Mol. Biol. 237, 75-86, 1994

A>Title: Complete sequence of the mitochondrial DNA of the chlorophyte alga Prototheca  
 A:Reference number: Z17373; MUID:94180393; PMID:9133522

A:Accession: T11918

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-78 <WOL>

A:Cross-references: EMBL:U02970; NID:9467843; PID:9467850; PIDN:AAD12637.1  
 A:Experimental source: strain 263-11

C:Genetics:

A:Gene: rps19  
 A:Genome: mitochondrion  
 C:Superfamily: Escherichia coli ribosomal protein S19  
 C:Keywords: mitochondrion; protein biosynthesis; ribosome

Query Match 11.0%; Score 48.5; DB 2; Length 78;  
 Best Local Similarity 33.3%; Pred. No. 2.6e+02;  
 Matches 13; Conservative 7; Mismatches 14; Indels 5; Gaps 2;

QY 30 KAMSGRRRLCCHRVSPNSTNLKGHHVRLCKPCCKLEPE 68  
 | | | : | : | : | | | : |  
 Db 23 KIMS-RRSAL----LPQVGRVTSIHNGRIFTCKISPE 56

RESULT 10:  
 F86885

30S ribosomal protein S14 [imported] - Lactococcus lactis subsp. lactis (strain IL140

C:Species: Lactococcus lactis subsp. lactis  
 C:Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 02-Aug-2002

C:Accession: F86885

R:Bohlin, A.; Winkler, P.; Manger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Eh  
 Genome Res. 11, 731-753, 2001

A>Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis  
 A:Reference number: A86625; MUID:21235186; PMID:11337471

A:Accession: F86885

A:Status: preliminary  
 A:Molecule type: DNA

A:Residues: 1-61 <STO>

A:Cross-references: GB:AE005176; PID:q12725139; PIDN:AK06184.1; GSPDB:GN00146  
 A:Experimental source: strain IL1403

C:Genetics:

A:Gene: rpsN  
 C:Superfamily: Escherichia coli ribosomal protein S14

Query Match 10.9%; Score 48; DB 2; Length 61;  
 Best Local Similarity 33.9%; Pred. No. 2.4e+02;  
 Matches 19; Conservative 5; Mismatches 22; Indels 10; Gaps 3;

QY 26 RRPKASGRRRFLCCHRVSPNSTNLKGHHVRLCKPCCKLEPRLMVVPGALPOV 81  
 | | | : | | | : | | | : | | | : |  
 Db 11 QRPKRSTQAYTR--CERGGRHSYVRK---FLCRICLR-----LAKGQLPGV 56

RESULT 11:  
 I59337

mammary transforming protein - mouse

C:Species: Mus musculus (house mouse)  
 C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999

C:Accession: I59337

R:Berz, T.K.; Guzman, R.C.; Miyamoto, S.; Panda, D.K.; Sasaki, M.; Hanyu, K.; Enami,  
 Proc. Natl. Acad. Sci. U.S.A. 91, 9789-9793, 1994

A>Title: Identification of a mammary transforming gene (MAT1) associated with mouse  
 A:Reference number: I59337; MUID:95024046; PMID:7937892

A:Accession: I59337

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-61 <RES>

A:Cross-references: GB:J31958; NID:9473909; PIDN:AA037665.1; PID:9473910  
 C:Genetics:

A:Gene: MAT1

Query Match 10.9%; Score 48; DB 2; Length 61;  
 Best Local Similarity 28.3%; Pred. No. 2.4e+02;  
 Matches 13; Conservative 7; Mismatches 16; Indels 10; Gaps 2;

QY 17 FSTFSTEGRRRPAPKASGR-----RRLCCHRVSPNSTNLKGHH 56  
 | | | : | | | : | | | : | | | : |  
 Db 13 FVVFSTINLSRPREWEKMPKSGRAKL---IQSPNKHISTNH 54

RESULT 12:  
 T47058

hypothetical protein [imported] - Yersinia pestis  
 C:Species: Yersinia pestis  
 C:Date: 17-Mar-2000 #sequence\_revision 17-Mar-2000 #text\_change 17-Mar-2000  
 C:Accession: T47058  
 R:Buchrieser, C.; Kusnlok, C.; Couve, E.; Frangeul, L.; Billaud, A.; Kunst, F.; Carr

submitted to the EMBL Data Library, October 1998

A:Description: DNA sequence of the 102 kbases unstable region of *Yersinia pestis*.

A:Reference number: Z24348

A:Accession: T47058

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-69 <BUC>

A:Cross-references: EMBL:AL031866; PIDN:CAA21401.1

A:Experimental source: strain 6/69

Query Match 10.9%; Score 48; DB 2; Length 69;

Best Local Similarity 25.5%; Pred. No. 2.6e+02;

Matches 14; Conservative 5; Mismatches 34; Indels 2; Gaps 1;

OY 16 CEFISTEGKRRPAKAMSGRRTRLCCHRVSPNSTNLKGHHVRLCKPCKLEPPER 70

DB 15 CLEALIESRLPPTPTSLQHCRCGKAIPEKRRQTLPG--VTCTDCQSLERK 67

### RESULT 13

GSF7 salivary glue protein sgs-7 - fruit fly (*Drosophila melanogaster*)

C:Species: *Drosophila melanogaster*

C:Date: 28-May-1986 #sequence\_revision 28-May-1986 #text\_change 16-Jul-1999

C:Accession: A03330

R:Garfinkel, M.D.; Pruitt, R.E.; Meyerowitz, E.M.

J. Mol. Biol. 168, 765-789, 1983

A:Title: DNA sequences, gene regulation and modular protein evolution in the *Drosophila*

A:Reference number: A92904; MUID:83294545; PMID:6411930

A:Accession: A03330

A:Molecule type: DNA

A:Residues: 1-74 <GAR>

A:Cross-references: GB:X01918; NID:g8581; PIDN:CAA5993.1; PID:g603988

C:Comment: This protein is produced by third-instar larvae.

C:Genetics:

A:Gene: sgs-7

A:Cross-references: FlyBase:FBgn0003377

A:Map position: 3L (68C)

A:Introns: 10/1

C:Superfamily: salivary glue protein

C:Keywords: salivary gland

Query Match

Best Local Similarity 10.9%; Score 48; DB 1; Length 74;

Matches 22; Conservative 18; Mismatches 26; Indels 16; Gaps 7;

OY 1 MRLVLSLCLILLCISISTEG--KRR---AKAWG--RRRLCCHRVSPNSTNL 52

DB 1 MKLIAVTLIACILILGFDLALGACQPCGPGKACTGCPKRPQLCQLISDIRNLQ 60

OY 53 KGHVRLCKPCKLEPPERLWV 74

DB 61 K---IRKC-VCG-EPQ---WMI 74

RESULT 14

T20201

hypothetical protein C54C8.9 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T20201

R:Dobson, R.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19237

A:Accession: T20201

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-79 <WIL>

A:Cross-references: EMBL:Z83102; PIDN:CAB05463.1; GSPDB:GN00019; CESP:C54C8.9

A:Experimental source: clone C54C8

C:Genetics:

A:Gene: CESP:C54C8.9

A:Map position: 1

Query Match 10.9%; Score 48; DB 2; Length 79;

Best Local Similarity 30.4%; Pred. No. 2.9e+02;

Matches 24; Conservative 11; Mismatches 26; Indels 18; Gaps 5;

OY 1 MRLVLSLCLILLCISISTEGKRRPAKAMSGRRTRLCCHR-----VSPNST 50

DB 1 MKLILFLFAIF--FEVILDSP-IEPFYSTGNSVPSHRHRIPLRGKREVNFQAD 56

OY 51 NL--KGHVR--LCKPCKL 65

DB 57 NVPEAGGRVRRYVGPPLKL 75

### RESULT 15

F90702

hypothetical protein ECs0590 [imported] - *Escherichia coli* (strain 0157:H7, substrain

C:Species: *Escherichia coli*

C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C:Accession: F90702

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.

gasawara, N.; Yasunaga, T.; Kunara, S.; Shida, T.; Hatvori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* 0157:H7 and g

A:Reference number: A9629; MUID:21156231; PMID:11258796

A:Accession: F90702

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-77 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA834013.1; PID:g13360048; GSPDB:GN00154

A:Experimental source: strain 0157:H7, substrain RMD 0509952

C:Genetics:

A:Gene: ECs0590

Query Match

Best Local Similarity 10.7%; Score 47.5; DB 2; Length 77;

Matches 15; Conservative 36.6%; Pred. No. 3.3e+02;

Matches 15; Conservative 5; Mismatches 16; Indels 5; Gaps 2;

OY 42 HRVSPNSTNLKG-HVRLCKPCKLEPPERLWVPGALPOV 81

DB 3 HRVSNMATFSLGKHVHVELCDLRLKLG---WSESGOAKI 39

Search completed: May 8, 2003, 21:48:09

Job time : 22 secs



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 8, 2003, 22:39:38 ; Search time 104 Seconds

(without alignments)  
967,876 Million cell updates/sec

Title: US-09-599-087b-5

Perfect score: 442  
Sequence: 1 MRLVLSLCLLCLLCLFSTF.....PCKLEPPRLWVPGALPGV 81

## Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 783854 seqs, 621352466 residues

Total number of hits satisfying chosen parameters: 238982

Minimum DB seq length: 25

Maximum DB seq length: 80

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Command line parameters:

-MODE=frame\_plus\_p2n.model -DEV=xlp  
-Q/cgn2\_1/USPTO.spool/US09599087.rnat.07052003.113853.7721/app\_query.fasta\_1.263  
-DB=Published.Applications.NA -OFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsun62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR.SCORE=pcr -THR.MAX=100  
-THR.MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pcr -NORM=ext -HEAPSIZ=500  
-MINLEN=25 -MAXLEN=80 -USER=US09599087.ecgn.1.1.93.ecgnat.07052003.113853.7721  
-NCPU=6 -ICPU=3 -NO.XLPXY -NO.MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT  
-DBPBLOCK=100 -LONGLOG -DEV.TIMEOUT=120 -MARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10  
-XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

Published.Applications.NA.\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCr\_NEM\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEM\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEM\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCrUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEM\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEM\_PUB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEM\_PUB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEM\_PUB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
C 1	44	10.0	50	US-10-126-617-7	Sequence 7, Appl1
C 2	44	10.0	50	US-10-126-617-8	Sequence 8, Appl1
C 3	43	9.7	69	US-10-159-428-1	Sequence 1, Appl1
C 4	43	9.7	78	US-10-204-200-6	Sequence 6, Appl1

C 5	43	9.7	78	US-10-204-200-7	Sequence 7, Appl1
C 6	43	9.7	78	US-09-950-844-5	Sequence 5, Appl1
C 7	43	9.7	78	US-09-950-844-6	Sequence 6, Appl1
C 8	43	9.7	78	US-09-835-699-5	Sequence 5, Appl1
C 9	43	9.7	78	US-09-835-699-6	Sequence 6, Appl1
C 10	42.5	9.6	59	US-10-006-593-89	Sequence 89, Appl1
C 11	42	9.5	50	US-10-137-788-6	Sequence 6, Appl1
C 12	41.5	9.4	46	US-09-827-864-19	Sequence 18, Appl1
C 13	41.5	9.4	54	US-09-294-093B-4383	Sequence 4383, Ap
C 14	41.5	9.4	60	US-09-902-941-1979	Sequence 1979, Ap
C 15	41.5	9.4	60	US-10-017-754-1979	Sequence 1979, Ap
C 16	41	9.3	60	US-09-954-692-13	Sequence 13, Appl
C 17	41	9.3	60	US-09-954-692-13	Sequence 13, Appl
C 18	41	9.3	80	US-09-440-829-17	Sequence 17, Appl
C 19	41	9.3	80	US-09-157-748-43	Sequence 43, Appl
C 20	40.5	9.2	75	US-09-864-761-28813	Sequence 16, Appl
C 21	40	9.0	31	US-09-773-599-16	Sequence 1511, Ap
C 22	40	9.0	31	US-09-801-274-1511	Sequence 1511, Ap
C 23	40	9.0	31	US-09-801-274-1512	Sequence 1512, Ap
C 24	40	9.0	46	US-09-940-244-227	Sequence 227, App
C 25	40	9.0	48	US-09-877-478-5348	Sequence 5348, Ap
C 26	40	9.0	59	US-10-025-380-976	Sequence 976, App
C 27	40	9.0	59	US-09-922-217-976	Sequence 976, App
C 28	40	9.0	59	US-09-833-263-976	Sequence 976, App
C 29	40	9.0	59	US-10-013-737-15	Sequence 15, Appl
C 30	40	9.0	60	US-09-902-941-1980	Sequence 1980, Ap
C 31	40	9.0	60	US-10-017-754-1980	Sequence 1980, Ap
C 32	40	9.0	63	US-09-923-876-1639	Sequence 1639, Ap
C 33	40	9.0	64	US-10-072-036-7	Sequence 7, Appl
C 34	40	9.0	75	US-09-979-999-15	Sequence 15, Appl
C 35	40	9.0	75	US-10-045-465-2	Sequence 2, Appl
C 36	40	9.0	78	US-09-864-761-21687	Sequence 21687, A
C 37	39.5	8.9	78	US-09-861-257-47	Sequence 47, Appl
C 38	39	8.8	25	US-10-215-112-13778	Sequence 13778, A
C 39	39	8.8	28	US-10-156-6344-13	Sequence 13, Appl
C 40	39	8.8	36	US-09-345-373-5	Sequence 5, Appl1
C 41	39	8.8	37	US-09-263-958-815	Sequence 815, App
C 42	39	8.8	40	US-09-263-958-815	Sequence 815, App
C 43	39	8.8	64	US-10-085-906-168	Sequence 168, App
C 44	39	8.8	65	US-09-747-003-12	Sequence 12, Appl
C 45	39	8.8	69	US-10-076-816-28	Sequence 28, Appl

## ALIGNMENTS

RESULT 1  
US-10-126-617-7/c  
; Sequence 7, Application US/10126617  
; Patent No. US20020168723A1  
; GENERAL INFORMATION:  
; APPLICANT: KAMBOJ, RAJENDER  
; APPLICANT: ELLIOTT, CANDACE E.  
; APPLICANT: NUTT, STEPHEN L.  
; TITLE OF INVENTION: KINASE-BINDING HUMAN CNS RECEPTORS OF THE EAA FAMILY  
; FILE REFERENCE: 016777/0471  
; CURRENT APPLICATION NUMBER: US/10/126,617  
; PRIOR FILING DATE: 2002-04-22  
; PRIOR APPLICATION NUMBER: 08/249,241  
; PRIOR FILING DATE: 1994-05-25  
; PRIOR APPLICATION NUMBER: 07/903,456  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 50  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-126-617-7  
ALIGNMENT Scores:  
Alignment No.: 566  
Score: 44.00  
Percent Similarity: 60.008  
Length: 50  
Matches: 7  
Conservative: 2



```

; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Human
US-10-204-200-7

Alignment Scores:
Pred. No.: 1.32e+03 Length: 78
Score: 43.00 Matches: 7
Percent Similarity: 83.33% Conservative: 3
Best Local Similarity: 58.33% Mismatches: 2
Query Match: 9.73% Indels: 0
DB: 9 Gaps: 0

US-09-599-087b-5 (1-81) x US-10-204-200-7 (1-78)
QY 9 Leuleucysilleleuleucyspheserillephe 20
||| |||:||||| ||| |||:|||||
Db 54 CTCTGCTGTGCTGCTGCTGTGTGTGAGACAGTCTTC 19

RESULT 6
US-09-950-844-5
; Sequence 5, Application US/09950844
; Patent No. US20020045594A1
; GENERAL INFORMATION:
; APPLICANT: Volkln, David B.
; APPLICANT: Evans, Robert K.
; APPLICANT: Uimer, Jeffrey B.
; APPLICANT: Caulfield, Michael J.
; TITLE OF INVENTION: POLYNUCLEOTIDE VACCINE FORMULATIONS
; FILE REFERENCE: 19907YIACA
; CURRENT APPLICATION NUMBER: US/09/950,844
; PRIOR FILING DATE: 1998-07-09
; PRIOR APPLICATION NUMBER: 09/112,655
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: 60/038,194
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Oligonucleotide
US-09-950-844-5

Alignment Scores:
Pred. No.: 1.32e+03 Length: 78
Score: 43.00 Matches: 7
Percent Similarity: 83.33% Conservative: 3
Best Local Similarity: 58.33% Mismatches: 2
Query Match: 9.73% Indels: 0
DB: 10 Gaps: 0

US-09-599-087b-5 (1-81) x US-09-950-844-5 (1-78)
QY 9 Leuleucysilleleuleucyspheserillephe 20
||| |||:||||| ||| |||:|||||
Db 29 CTCTGCTGTGCTGCTGCTGTGTGTGAGACAGTCTTC 64

RESULT 7
US-09-950-844-6/c
; Sequence 6, Application US/09950844
; Patent No. US20020045594A1
; GENERAL INFORMATION:
; APPLICANT: Volkln, David B.
; APPLICANT: Evans, Robert K.
; APPLICANT: Uimer, Jeffrey B.

; APPLICANT: Caulfield, Michael J.
; TITLE OF INVENTION: POLYNUCLEOTIDE VACCINE FORMULATIONS
; FILE REFERENCE: 19907YIACA
; CURRENT APPLICATION NUMBER: US/09/950,844
; PRIOR FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 09/112,655
; PRIOR FILING DATE: 1998-07-09
; PRIOR APPLICATION NUMBER: 09/023,834
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: 60/038,194
; PRIOR FILING DATE: 1997-02-14
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Oligonucleotide
US-09-950-844-6

Alignment Scores:
Pred. No.: 1.32e+03 Length: 78
Score: 43.00 Matches: 7
Percent Similarity: 83.33% Conservative: 3
Best Local Similarity: 58.33% Mismatches: 2
Query Match: 9.73% Indels: 0
DB: 10 Gaps: 0

US-09-599-087b-5 (1-81) x US-09-950-844-6 (1-78)
QY 9 Leuleucysilleleuleucyspheserillephe 20
||| |||:||||| ||| |||:|||||
Db 54 CTCTGCTGTGCTGCTGCTGTGTGTGAGACAGTCTTC 19

RESULT 8
US-09-835-699-5
; Sequence 5, Application US/09835699
; Patent No. US20020147167A1
; GENERAL INFORMATION:
; APPLICANT: Armstrong, Marcy E.
; APPLICANT: Lewis, John A.
; APPLICANT: Liu, Margaret A.
; APPLICANT: McClements, William L.
; TITLE OF INVENTION: A POLYNUCLEOTIDE HERPES VIRUS
; VACCINE
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: P.O. Box 2000, 126 E. Lincoln Ave.
; CITY: Rahway
; STATE: NJ
; COUNTRY: USA
; ZIP: 07065-0907
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/835,699
; FILING DATE: 17-Aug-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCF/US97/17262
; FILING DATE: 26-SEP-1997
; APPLICATION NUMBER: US 08/720,758
; FILING DATE: 01-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Hand, J. Mark
; REGISTRATION NUMBER: 36,545
; REFERENCE/DOCKET NUMBER: 19258CC
```



```
Sequence 6, Application US/10137788
; Publication No. US20030039636A1
; GENERAL INFORMATION:
; APPLICANT: Lebouich, Philippe
; APPLICANT: Westerman, Karen
; TITLE OF INVENTION: NOVEL SELF-INACTIVATING (SIN) LENTIVIRAL VECTORS
; FILE REFERENCE: 101-023
; CURRENT APPLICATION NUMBER: US/10/137,788
; CURRENT FILING DATE: 2002-05-01
; PRIOR APPLICATION NUMBER: US 60/288,042
; PRIOR FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 50
; TYPE: DNA
; ORGANISM: Human Immunodeficiency Virus type 1
US-10-137-788-6

Alignment Scores:
Pred. No.: 1.01e+03 Length: 50
Score: 42.00 Matches: 7
Percent Similarity: 73.33% Conservative: 4
Best Local Similarity: 46.67% Mismatches: 4
Query Match: 9.50% Indels: 0
DB: Gaps: 0

US-09-599-087B-5 (1-81) x US-10-137-788-6 (1-50)

Oy 4 LeuValleuserleucystilleuLeuLeucyspheser 18
Db 4 CTGCGCTGAGTGGCTCAATGCTGTGCTTTTGTGTCTCG 48

RESULT 12
US-09-827-864-19/c
; Sequence 19, Application US/09827864
; Patent No. US20020009458A1
; GENERAL INFORMATION:
; APPLICANT: COLAU, DIDIER
; ROOS, JOEL
; TITLE OF INVENTION: RECOMBINANT DNA SEQUENCES, MOLECULES,
; VECTORS AND VACCINES FOR FELINE
; CALICIVIRUS DISEASE AND METHODS FOR
; PRODUCING AND USING SAME.
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MCDERMOTT, WILL & EMERY
; STREET: 1850 K STREET, N.W., SUITE 500
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/827,864
; FILING DATE: 06-Apr-2001
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: GADIANO, WILHELM F
; REGISTRATION NUMBER: 37,136
; REFERENCE/DOCKET NUMBER: 37/12-213
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 778-8373
; TELEFAX: (202) 778-8335
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
```

```
TOPOLOGY: linear
; MOLECULE TYPE: nucleic acid
; SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-09-827-864-19

Alignment Scores:
Pred. No.: 1.05e+03 Length: 46
Score: 41.50 Matches: 8
Percent Similarity: 64.29% Conservative: 1
Best Local Similarity: 57.14% Mismatches: 4
Query Match: 9.39% Indels: 1
DB: Gaps: 1

US-09-599-087B-5 (1-81) x US-09-827-864-19 (1-46)

Oy 60 CysLysProCysLysLeuGluProGluProArgLeuTyrVal 73
Db 42 TGTCACTGTGCAAGCTGGAATGACGCTCTCT--TGTTTA 4

RESULT 13
US-09-294-093B-4383/c
; Sequence 4383, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Laligudi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; CURRENT FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 4383
; LENGTH: 54
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID NO. US20010051335A1 700353949H1
; NAME/KEY: unsure
; LOCATION: 34
; OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-4383

Alignment Scores:
Pred. No.: 1.28e+03 Length: 54
Score: 41.50 Matches: 8
Percent Similarity: 69.23% Conservative: 1
Best Local Similarity: 61.54% Mismatches: 1
Query Match: 9.39% Indels: 3
DB: Gaps: 1

US-09-599-087B-5 (1-81) x US-09-294-093B-4383 (1-54)

Oy 34 GlyArgArgThrArgLeuCysGlyShsArgValProSer 46
Db 52 GGACGGAGG-----TGCTGCCATNTATACCAAGT 23

RESULT 14
US-09-902-941-1979
; Sequence 1979, Application US/09902941
; Patent No. US20020172952A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Marnerakis, Margarita
; APPLICANT: Carter, Darick
```

Search completed: May 8, 2003, 23:29:51  
Job time : 104 secs

```

; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.478C17
; CURRENT APPLICATION NUMBER: US/09/902,941
; CURRENT FILING DATE: 2001-07-10
; NUMBER OF SEQ ID NOS: 2002
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1979
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-902-941-1979
```

```

Alignment Scores:
Pred. No.: 1.46e+03 Length: 60
Score: 41.50 Matches: 9
Percent Similarity: 61.11% Conservative: 2
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 9.39% Indels: 1
DB: 9 Gaps: 1
```

US-09-599-087b-5 (1-81) x US-09-902-941-1979 (1-60)

```
OY 34 GLYArgArgThrArgLeu---CysCysHisArgValProSerProAsnSerThr 50
|||:::||||| ||||| ||||| |||||
DB 1 GGGGCTCAGCGCAGCTGGGGTCTGTGGGGTATCCGAGTCCAGAGCACC 54
```

#### RESULT 15

```
US-10-017-754-1979
; Sequence 1979, Application US/10017754
; Publication No. US20030054363A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tonglong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Mainerakis, Margarita
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.478C18
; CURRENT APPLICATION NUMBER: US/10/017,754
; CURRENT FILING DATE: 2001-10-29
; NUMBER OF SEQ ID NOS: 2004
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1979
; LENGTH: 60
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-017-754-1979
```

```

Alignment Scores:
Pred. No.: 1.46e+03 Length: 60
Score: 41.50 Matches: 9
Percent Similarity: 61.11% Conservative: 2
Best Local Similarity: 50.00% Mismatches: 6
Query Match: 9.39% Indels: 1
DB: 9 Gaps: 1
```

US-09-599-087b-5 (1-81) x US-10-017-754-1979 (1-60)

```
OY 34 GLYArgArgThrArgLeu---CysCysHisArgValProSerProAsnSerThr 50
|||:::||||| ||||| ||||| |||||
DB 1 GGGGCTCAGCGCAGCTGGGGTCTGTGGGGTATCCGAGTCCAGAGCACC 54
```

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 8, 2003, 22:37:43 ; Search time 62 Seconds  
(without alignments)

400.658 Million cell updates/sec

Title: US-09-599-087b-5

Perfect score: 442

Sequence: 1 MRLVLSLLCILLICFSIF.....PCKLEPERLWVPGALPGV 81

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 441362 segs, 15338381 residues

Total number of hits satisfying chosen parameters: 322676

Minimum DB seq length: 25

Maximum DB seq length: 80

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL-framer\_p2n.model -DEV-xlp  
-Q/cgn2\_1/USPTO.spool/US05959087/rnuc.07052003\_113853\_7693/app\_query.fasta\_1.263  
-DB-Issued-Patents\_NA -OFMT-fastap -SUFFIX-rn1 -MINMATCH-0.1 -LOOPEL-0  
-LOPEXT-0 -UNITS-bits -START-1 -END-1 -MATRIX-biosum62 -TRANS-human40.cdi  
-LIST-45 -DOALIGN-pto -THR SCORE-pct -THR MAX-100 -THR MIN-0 -ALIGN-15  
-MODE-LOCAL -CUTPOINT-pto -NORM-ext HEADSIZE-500 -MINLEN-25 -MAXLEN-80  
-USER-US05959087@cgn2\_1.36@rnuc.07052003\_113853\_7693 -NCPU-6 -ICPU-3  
-NO\_XLPHY -NO\_MAP -LARGEOUTERY -NEG\_SCORES-0 -WAIT -DSPBLOCK-100 -LONGLOG  
-DEV\_TIMEOUT-120 -WARN\_TIMEOUT-30 -THREADS-1 -XGAPOP-10 -XGAPEXT-0.5 -FGAPOP-6  
-FAPEXT-7 -YGAPOP-10 -YGAPEXT-0.5 -DELOP-6 -DELEXT-7

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/lna/3A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/lna/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/lna/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/lna/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/lna/PCTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/lna/backtiltestl.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53	12.0	79	1	US-08-170-290A-19
2	46.5	10.5	48	2	US-08-350-260A-289
3	46	10.4	72	1	US-08-441-430-22
4	45.5	10.3	69	1	US-08-554-612C-37
5	45	10.2	39	2	US-08-350-260A-233
6	44	10.0	42	1	US-08-642-255-23
7	44	10.0	50	1	US-07-803-456-7
8	44	10.0	50	1	US-07-803-456-8
9	44	10.0	50	4	US-08-249-241-7
10	44	10.0	50	4	US-08-249-241-8
11	44	10.0	51	2	US-08-350-260A-287
12	44	10.0	51	2	US-08-350-260A-290

13	43	9.7	72	6	5466668-7	Patent No. 5466668
14	43	9.7	78	1	US-08-338-992B-5	Sequence 5, Appl1
15	43	9.7	78	1	US-08-338-992B-6	Sequence 5, Appl1
16	43	9.7	78	4	US-09-010-733-5	Sequence 6, Appl1
17	43	9.7	78	4	US-09-010-733-6	Sequence 6, Appl1
18	43	9.7	78	5	PCT-US95-09057-5	Sequence 6, Appl1
19	43	9.7	78	5	PCT-US95-09057-6	Sequence 6, Appl1
20	43	9.7	79	1	US-08-571-983-2	Sequence 2, Appl1
21	43	9.7	79	5	PCT-US94-08111-2	Sequence 2, Appl1
22	42	9.5	30	1	US-08-338-471D-4	Sequence 4, Appl1
23	42	9.5	69	1	US-07-977-284A-207	Sequence 207, App
24	42	9.5	69	2	US-08-256-426B-207	Sequence 19, App
25	41.5	9.4	46	4	US-08-872-056-19	Sequence 112, App
26	41.5	9.4	60	2	US-08-465-380-112	Sequence 27, App
27	41.5	9.4	60	2	US-08-480-478-27	Sequence 112, App
28	41.5	9.4	60	2	US-08-486-397-112	Sequence 112, App
29	41.5	9.4	60	2	US-08-486-399-112	Sequence 112, App
30	41.5	9.4	60	2	US-08-461-965-112	Sequence 112, App
31	41.5	9.4	60	2	US-08-326-110A-27	Sequence 27, Appl
32	41.5	9.4	60	2	US-08-634-641-112	Sequence 112, App
33	41.5	9.4	60	3	US-09-249-471-112	Sequence 112, App
34	41.5	9.4	60	3	US-09-249-472-112	Sequence 112, App
35	41.5	9.4	60	3	US-09-249-451-112	Sequence 112, App
36	41.5	9.4	60	3	US-08-809-455-112	Sequence 112, App
37	41.5	9.4	60	3	US-09-249-461-112	Sequence 112, App
38	41.5	9.4	60	3	US-09-249-448-112	Sequence 112, App
39	41	9.3	30	1	US-07-869-931-27	Sequence 27, App
40	41	9.3	30	1	US-07-855-417A-27	Sequence 27, Appl
41	41	9.3	60	4	US-09-339-913B-13	Sequence 13, Appl
42	41	9.3	60	4	US-09-339-904A-13	Sequence 13, Appl
43	41	9.3	60	4	US-08-769-062B-13	Sequence 13, Appl
44	41	9.3	60	4	US-09-344-002B-13	Sequence 13, Appl
45	41	9.3	60	4	US-09-559-565C-13	Sequence 13, Appl

#### ALIGNMENTS

RESULT 1  
US-08-170-290A-19  
; Sequence 19, Application US/08170290A  
; Patent No. 5702931  
; GENERAL INFORMATION:  
; APPLICANT: Andrews, William H.  
; APPLICANT: Morser, Michael J.  
; APPLICANT: Ziehlender, Laura R.  
; TITLE OF INVENTION: No. 5702931el Mutagenesis Methods and  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: James M. Heslin  
; STREET: 379 Lytton Ave.  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94301  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/170,290A  
; FILING DATE: 28-DEC-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/05573  
; FILING DATE: 01-JUL-1992  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/724,237  
; FILING DATE: 01-JUL-1991  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:

NAME: Heslin, James M.  
REGISTRATION NUMBER: 29,541  
REFERENCE/DOCKET NUMBER: 11972-58-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 79 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-170-290A-19

Alignment Scores:  
Pred. No.: 114 Length: 79  
Score: 53.00 Matches: 9  
Percent Similarity: 55.00% Conservative: 2  
Best Local Similarity: 45.00% Mismatches: 9  
Query Match: 11.99% Indels: 0  
DB: 1 Gaps: 0

US-09-599-087B-5 (1-81) x US-08-170-290A-19 (1-79)

QY 29 AlAlysAlarPserGlyArgGlyThrArgLeuCyScyShsArgValProSerProAsn 48  
Db 15 GCGTCGGCGTGGCGGGGCTGCCAGAGCCGCTGCGCCACCGCTTCACCTTGCCGCGAGT 74

RESULT 2  
US-08-350-260A-289  
Sequence 289, Application US/08350260A  
Patent No. 5962255  
GENERAL INFORMATION:  
APPLICANT: Winter, Gregory Paul  
APPLICANT: Griffiths, Andrew David  
APPLICANT: Williams, Samuel Cameron  
APPLICANT: Waterhouse, Peter  
APPLICANT: Nissim, Anava  
APPLICANT: Johnson, Kevin Stuart  
TITLE OF INVENTION: Methods for producing members of specific  
TITLE OF INVENTION: binding pairs  
NUMBER OF SEQUENCES: 602  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David W. Clough  
STREET: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606-6402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/350,260A  
FILING DATE: 05-DEC-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9110549.4  
FILING DATE: 15-MAY-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9206318.9  
FILING DATE: 24-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB91/01134  
FILING DATE: 10-JUL-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB92/00883  
FILING DATE: 15-MAY-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB93/00605  
FILING DATE: 24-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/150,002  
FILING DATE: 31-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/307,619  
FILING DATE: 16-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Clough, David W  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 28111/32372  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
INFORMATION FOR SEQ ID NO: 289:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 48 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-350-260A-289

Alignment Scores:  
Pred. No.: 320 Length: 48  
Score: 46.50 Matches: 9  
Percent Similarity: 66.67% Conservative: 3  
Best Local Similarity: 50.00% Mismatches: 3  
Query Match: 10.52% Indels: 3  
DB: 2 Gaps: 1

US-09-599-087B-5 (1-81) x US-08-350-260A-289 (1-48)

QY 32 TTPSerGlyArgGlyThrArgLeuCyScyShsArgValProSerProAsnSer 49  
Db 3 TGGTCCCTCCGCGAATACAGTCTGCTGT-----CCACGCGCTGACAGT 47

RESULT 3  
US-08-441-430-22/C  
Sequence 22, Application US/08441430  
Patent No. 5681942  
GENERAL INFORMATION:  
APPLICANT: Buchwald, Manuel  
APPLICANT: Strathdee, Craig A.  
APPLICANT: Wevrick, Rachel  
APPLICANT: Mathew, Christopher George Porter  
TITLE OF INVENTION: Fanconi Anemia Type C Gene  
NUMBER OF SEQUENCES: 73  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Richard J. Polley, Esq.  
ADDRESSEE: Klarquist, Spateman, Campbell, Leigh &  
ADDRESSEE: Whinston, LLP  
STREET: 121 S.W. Salmon, Suite 1600  
CITY: Portland  
STATE: Oregon  
COUNTRY: U.S.A.  
ZIP: 97204  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Disk, 3+-inch  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: MS DOS  
SOFTWARE: WordPerfect 5.1/ASCII Text file  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/441,430  
FILING DATE: May 15, 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.S. 07/876,285  
FILING DATE: April 29, 1992  
APPLICATION NUMBER: U.S. 07/918,313  
FILING DATE: July 21, 1992  
APPLICATION NUMBER: U.S. 08/003,963  
FILING DATE: January 15, 1993



ATTORNEY/AGENT INFORMATION:  
NAME: Richard J. Polley, Esq.  
REGISTRATION NUMBER: 28,107  
REFERENCE/DOCKET NUMBER: 3812-42824  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (503) 226-7391  
TELEFAX: (503) 228-9446  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 72 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Double stranded  
TOPOLOGY: Linear  
MOLECULE TYPE: Genomic DNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
US-08-441-430-22

Alignment Scores:  
Pred. No.: 642  
Score: 46.00  
Percent Similarity: 76.92%  
Best Local Similarity: 53.85%  
Query Match: 10.41%  
DB: 1  
US-08-599-087B-5 (1-81) x US-08-441-430-22 (1-72)

QY 41 CysH1sArgYalPProSerPProAsnSerTHAsnLeuLys 53  
|||||::: ::||| ||| ||::|||  
Db 71 TGCACAGATGAGAAATCCAAAGAGCATGACATTAAAG 33

RESULT 4:  
US-08-554-612C-37  
Sequence 37, Application US/08554612C  
Patent No. 5747660  
GENERAL INFORMATION:  
APPLICANT: Orllick, David  
TITLE OF INVENTION: PROTAGLANDIN F2 RECEPTOR REGULATORY  
TITLE OF INVENTION: PROTEIN AND THERAPEUTIC USES  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 2200 Sand Hill Road, Suite 100  
CITY: Menlo Park  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 94025  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/554,612C  
FILING DATE: No. 5747660ember 6, 1995  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Sherwood, Pamela  
REGISTRATION NUMBER: 36,677  
REFERENCE/DOCKET NUMBER: 06519/004001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 322-5070  
TELEFAX: (415) 854-0875  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 69 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA

```

US-08-554-612C-37
Alignment Scores:
Pred. No.: 691
Score: 45.50
Percent Similarity: 54.55%
Best Local Similarity: 45.45%
Query Match: 10.25%
Indels: 3
Gaps: 1

US-09-599-087B-5 (1-81) x US-08-554-612C-37 (1-69)
OY 25 LysArgProIla-----LysAlaTrpSerGlyArgArgThrArgLeuGysCys 41
Db 3 AGGGAGAGATGCTCCCGCCGCGAGACATGGGGCGCCGCGCCGAGAGCCCTGCTGC 62
OY 42 HisArg 43
|||
Db 63 TGGCGC 68

RESULT 5
US-08-350-260A-233
Sequence 233, Application US/08350260A
Patent No. 5962255
GENERAL INFORMATION:
APPLICANT: Winter, Gregory Paul
APPLICANT: Griffiths, Andrew David
APPLICANT: Williams, Samuel Cameron
APPLICANT: Waterhouse, Peter
APPLICANT: Nieslin, Ahuva
APPLICANT: Johnson, Kevin Stuart
TITLE OF INVENTION: Methods for producing members of specific
TITLE OF INVENTION: binding pairs
NUMBER OF SEQUENCES: 602
CORRESPONDENCE ADDRESS:
ADDRESSEE: David W. Clough
STREET: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
City: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/350,260A
FILING DATE: 05-DEC-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9110549.4
FILING DATE: 15-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9206318.9
FILING DATE: 24-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB91/01134
FILING DATE: 10-JUL-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB92/00883
FILING DATE: 15-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB93/00605
FILING DATE: 24-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/150,002
FILING DATE: 31-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/307,619
FILING DATE: 16-SEP-1994

```

ATTORNEY/AGENT INFORMATION:  
NAME: Clough, David W  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 28111/32372  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
INFORMATION FOR SEQ ID NO: 233:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 39 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-350-260A-233

Alignment Scores:  
Score: 357  
Length: 39  
Percent Similarity: 45.00  
Matches: 8  
Best Local Similarity: 66.67%  
Conservative: 0  
Mismatch: 4  
Query Match: 10.188  
Indels: 0  
Gaps: 0

US-09-599-087B-5 (1-81) x US-08-350-260A-233 (1-39)  
Qy 32 TTPSeGlyArGArGhArGLeuGcYSHsArG 43  
DB 3 TGTGCTCCGCGCGAATACAGGCTGCTGCATAGG 38

RESULT 6  
US-08-642-255-23  
Sequence 23, Application US/08642255  
Patent No. 5773249  
GENERAL INFORMATION:  
APPLICANT: CAPPELLO, Joseph  
TITLE OF INVENTION: FERRARI, Franco A.  
TITLE OF INVENTION: High Molecular Weight Collagen-Like  
NUMBER OF SEQUENCES: 135  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HONBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Elbarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/642,255  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: ROWLAND, Bettam I.  
REGISTRATION NUMBER: 20,015  
REFERENCE/DOCKET NUMBER: A55556-3/BIR  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 494-8700  
TELEFAX: (415) 494-8771  
TELEX: 910 277299 FHT UR  
INFORMATION FOR SEQ ID NO: 23:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 42 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: other nucleic acid  
DESCRIPTION: /desc = "synthetic"  
US-08-642-255-23  
Alignment Scores:

Pred. No.: 516  
Score: 44.00  
Percent Similarity: 75.00%  
Best Local Similarity: 58.33%  
Query Match: 9.95%  
DB: 1  
Gaps: 0

US-09-599-087B-5 (1-81) x US-08-642-255-23 (1-42)  
Qy 27 ArgProAlaLysAlaTrpSerGlyArGArGhArG 38  
DB 4 AGGCCAGAGAGCTGCTGCGGCTCAGGCGAGCCAGG 39

RESULT 7  
US-07-903-456-7/C  
Sequence 7, Application US/07903456  
Patent No. 5574144  
GENERAL INFORMATION:  
APPLICANT: KAMBOJ, Rajender  
APPLICANT: ELLIOTT, Candace  
APPLICANT: NOTT, Stephen  
TITLE OF INVENTION: KAINATE-BINDING HUMAN CNS RECEPTORS OF  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/903,456  
FILING DATE: 19920624  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 16777/183/ALLE  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-9300  
TELEFAX: (703) 683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-07-903-456-7

Alignment Scores:  
Pred. No.: 657  
Score: 44.00  
Percent Similarity: 60.00%  
Best Local Similarity: 46.67%  
Query Match: 9.95%  
DB: 1  
Gaps: 0

US-09-599-087B-5 (1-81) x US-07-903-456-7 (1-50)  
Qy 34 GlyArGArGhArGLeuGcYSHsArGValProSerProAsn 48  
DB 46 GGCATGAGCTCAGAACCTTCTGCTGATGAGAGCTCAACTCAAC 2

RESULT 8  
US-07-903-456-8/C  
Sequence 8, Application US/07903456



TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 50 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-249-241-8

Alignment Scores:  
Pred. No.: 657 Length: 50  
Score: 44.00 Matches: 7  
Percent Similarity: 60.00% Conservative: 2  
Best Local Similarity: 46.67% Mismatches: 6  
Query Match: 9.95% Indels: 0  
Gaps: 0

US-09-599-087b-5 (1-81) x US-08-249-241-8 (1-50)

QY 34 GtAgtAgtTtArGLeucScyShIsArGyAlProSerProAsn 48  
Db 46 GGCATGAGCTCAGACCTTGTCATGAGAGCTCCACTCCAAAC 2

RESULT 11  
US-08-350-260A-287  
Sequence 287, Application US/08350260A  
Patent No. 5962255

GENERAL INFORMATION:  
APPLICANT: Winter, Gregory Paul  
APPLICANT: Griffiths, Andrew David  
APPLICANT: Williams, Samuel Cameron  
APPLICANT: Waterhouse, Peter  
APPLICANT: Nissim, Ahuva  
APPLICANT: Johnson, Kevin Stuart  
APPLICANT: Smith, Andrew John Hammond  
TITLE OF INVENTION: Methods for producing members of specific  
TITLE OF INVENTION: binding pairs  
NUMBER OF SEQUENCES: 602

CORRESPONDENCE ADDRESS:  
ADDRESSEE: David W. Clough  
STREET: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606-6402

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/350,260A  
FILING DATE: 05-DEC-1994  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9110549.4  
FILING DATE: 15-MAY-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9206318.9  
FILING DATE: 24-MAR-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB91/01134  
FILING DATE: 10-JUL-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB92/00883  
FILING DATE: 15-MAY-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB93/00605  
FILING DATE: 24-MAR-1993

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/150,002  
FILING DATE: 31-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/307,619  
FILING DATE: 16-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Clough, David W  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 28111/32372  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
INFORMATION FOR SEQ ID NO: 287:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 51 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-350-260A-287

Alignment Scores:  
Pred. No.: 676 Length: 51  
Score: 44.00 Matches: 9  
Percent Similarity: 62.50% Conservative: 1  
Best Local Similarity: 56.25% Mismatches: 4  
Query Match: 9.95% Indels: 2  
Gaps: 1

US-09-599-087b-5 (1-81) x US-08-350-260A-287 (1-51)

QY 32 TtPserGtArGArGtThArGLeucScyShIsArGyAlPro 45  
Db 3 TGGTCCCTCCGCCGAAATGCAAGTTCCTCCACTCCATGCTGCC 50

RESULT 12  
US-08-350-260A-290  
Sequence 290, Application US/08350260A  
Patent No. 5962255

GENERAL INFORMATION:  
APPLICANT: Winter, Gregory Paul  
APPLICANT: Griffiths, Andrew David  
APPLICANT: Williams, Samuel Cameron  
APPLICANT: Waterhouse, Peter  
APPLICANT: Nissim, Ahuva  
APPLICANT: Johnson, Kevin Stuart  
APPLICANT: Smith, Andrew John Hammond  
TITLE OF INVENTION: Methods for producing members of specific  
TITLE OF INVENTION: binding pairs  
NUMBER OF SEQUENCES: 602

CORRESPONDENCE ADDRESS:  
ADDRESSEE: David W. Clough  
STREET: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606-6402

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/350,260A  
FILING DATE: 05-DEC-1994  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9110549.4  
FILING DATE: 15-MAY-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9206318.9  
FILING DATE: 24-MAR-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/GB91/01134  
FILING DATE: 10-JUL-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB92/00883  
FILING DATE: 15-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB93/00605  
FILING DATE: 24-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/150,002  
FILING DATE: 31-MAR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/307,619  
FILING DATE: 16-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: CLOUGH, David W  
REGISTRATION NUMBER: 36,107  
REFERENCE/DOCKET NUMBER: 28111/32372  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
INFORMATION FOR SEQ ID NO: 290:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 51 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-350-260A-290

Alignment Scores:  
Pred. No.: 676 Length: 51  
Score: 44.00 Matches: 9  
Percent Similarity: 61.11% Conservative: 2  
Best Local Similarity: 50.00% Mismatches: 5  
Query Match: 9.95% Indels: 2  
DB: 2 Gaps: 1

US-09-599-087b-5 (1-81) x US-08-350-260A-290 (1-51)  
QY 32 TTPserGIyArGaTgThArGleuCysCyshIstArGyAlProserProasnsr 49  
Db 3 TGTGCTGTCGCGCGAATACMNNAGTGCTGC-----TGTCCACGCGCTGACAGT 50

RESULT 13  
5466668-7  
Patent No. 5466668  
APPLICANT: GLASER, CHARLES B.; MORSE, MICHAEL J.; LIGHT,  
DAVID R.  
TITLE OF INVENTION: SUPERIOR THROMBOMODULIN ANALOGS FOR  
PHARMACEUTICAL USE  
NUMBER OF SEQUENCES: 57  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/155,346  
FILING DATE: 22-NOV-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 568,456  
FILING DATE: 15-AUG-1990  
APPLICATION NUMBER: 506,325  
FILING DATE: 09-APR-1990  
APPLICATION NUMBER: 406,941  
FILING DATE: 13-SEP-1989  
APPLICATION NUMBER: 345,374  
FILING DATE: 28-APR-1989  
SEQ ID NO: 7:  
LENGTH: 72  
5466668-7

Alignment Scores:  
Pred. No.: 1,42e+03 Length: 72  
Score: 43.00 Matches: 7  
Percent Similarity: 83.33% Conservative: 3  
Best Local Similarity: 58.33% Mismatches: 2  
Query Match: 9.73% Indels: 0  
DB: 6 Gaps: 0

US-09-599-087b-5 (1-81) x 5466668-7 (1-72)  
QY 9 LeuLeuCysIleLeuLeuLeuCysPhseSerIlePhe 20  
Db 22 CTCGTGCTGTGTGCTGCTGTGTGTGAGACATCTTC 57

RESULT 14  
US-08-338-992B-5  
Sequence 5, Application US/08338992B  
Patent No. 5736524  
GENERAL INFORMATION:  
APPLICANT: CONTENT, JEAN  
APPLICANT: HUYGEN, KRIS  
APPLICANT: LIOU, MARGARET A.  
APPLICANT: MONTGOMERY, DONNA  
APPLICANT: ULMER, JEFFREY  
TITLE OF INVENTION: A POLYNUCLEOTIDE TUBERCULOSIS VACCINE  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MICHAEL D. YABLONSKY - MERCK & CO., INC.  
STREET: 126 E. LINCOLN AVE., - P.O. BOX 2000  
CITY: RAHWAY  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07065-0900  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/338,992B  
FILING DATE: 14-NOV-1994  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: YABLONSKY, MICHAEL D.  
REGISTRATION NUMBER: 40,407  
REFERENCE/DOCKET NUMBER: 19342  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 908-594-4678  
TELEFAX: 908-594-4720  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 78 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-338-992B-5

Alignment Scores:  
Pred. No.: 1,59e+03 Length: 78  
Score: 43.00 Matches: 7  
Percent Similarity: 83.33% Conservative: 3  
Best Local Similarity: 58.33% Mismatches: 2  
Query Match: 9.73% Indels: 0  
DB: 1 Gaps: 0

US-09-599-087b-5 (1-81) x US-08-338-992B-5 (1-78)  
QY 9 LeuLeuCysIleLeuLeuLeuCysPhseSerIlePhe 20  
Db 29 CTCGTGCTGTGTGCTGCTGTGTGTGAGACATCTTC 64

RESULT 15  
US-08-338-992B-6/c  
Sequence 6, Application US/08338992B  
Patent No. 5736524  
GENERAL INFORMATION:  
APPLICANT: CONTENT, JEAN  
APPLICANT: HUYGEN, KRIS  
APPLICANT: LIOU, MARGARET A.

Fri May 9 16:38:51 2003

us-09-599-087b-5.rn1

Page 8

```

1  APPLICANT: MONTGOMERY, DONNA
2  APPLICANT: ULMER, JEFFREY
3  TITLE OF INVENTION: A POLYNUCLEOTIDE TUBERCULOSIS VACCINE
4  NUMBER OF SEQUENCES: 35
5  CORRESPONDENCE ADDRESS:
6  ADDRESSEE: MICHAEL D. YABLONSKY - MERCK & CO., INC.
7  STREET: 126 E. LINCOLN AVE., - P.O. BOX 2000
8  CITY: RAHWAY
9  STATE: NEW JERSEY
10 COUNTRY: USA
11 ZIP: 07065-0900
12
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: Floppy disk
15 COMPUTER: IBM PC compatible
16 OPERATING SYSTEM: PC-DOS/MS-DOS
17 SOFTWARE: PatentIn Release #1.0, Version #1.30
18
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER: US/08/338,992B
21 FILING DATE: 14-NOV-1994
22 CLASSIFICATION: 514
23
24 ATTORNEY/AGENT INFORMATION:
25 NAME: YABLONSKY, MICHAEL D.
26 REGISTRATION NUMBER: 40,407
27 REFERENCE/DOCKET NUMBER: 19342
28 TELECOMMUNICATION INFORMATION:
29 TELEPHONE: 908-594-4678
30 TELEFAX: 908-594-4720
31
32 INFORMATION FOR SEQ. ID NO.: 6:
33 SEQUENCE CHARACTERISTICS:
34 LENGTH: 78 base pairs
35 TYPE: nucleic acid
36 STRANDEDNESS: single
37 TOPOLOGY: linear
38
39 MOLECULE TYPE: DNA (genomic)
40
41 OS-08-338-992B-6

```

Alignment Scores:	
Pred. No.:	1.59e+03
Score:	43.00
Percent Similarity:	83.93%
Best Local Similarity:	58.53%
Query Match:	5.73%
DB:	1
	Gaps: 0
	Indels: 0
	Mismatches: 2
	Conservative: 3
	Matches: 78
	Length: 78

US-09-599-087B-5 (1-81) x US-08-338-992B-6 (1-78)

QY 9 LeuLeuCysIleLeuLeuLeuCysPheSerIlePhe 20  
||| |||:::||||| |||::: |||  
Db 54 CTCTGCTGTGTGCTGCTGTGTGTGGAGCAGTCTTC 19

Search completed: May 8, 2003, 23:27:54  
Job time : 63 secs

Page 8

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 8, 2003, 21:53:08 ; Search time 207 Seconds  
(without alignments)

881.217 Million cell updates/sec

Title: US-09-599-087B-5

Sequence: 1 MRLVLSLLCLLCISIF.....PCKLEPRRWVPCALQV 81

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 710428

Minimum DB seq length: 25  
Maximum DB seq length: 80

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters: -DEV-xlp  
-MODE-frame+ p2n.model -US09599087/runat\_07052003\_113851.7658/app\_query.fasta\_1.263  
-DB-N.Geneseq 101002 -OFMT-fastap -SUFFIX-rng -MINMATCH-0.1 -IOOPT-0  
-LOOPT-0 -UNITS-bits -STAR-1 -END-1 -MATRIX-biosum62 -TRANS-human40.cdi  
-LIST-45 -DOCALLIGN-200 -THR-SCORE-pct -THR-MAX-100 -THR-MIN-0 -ALIGN-15  
-MODE-LOCAL -OUTFMT-pto -NORM-ext -HEADSIZE-500 -MINLEN-25 -MAXLEN-80  
-USER-US09599087-@CGN\_1\_1.208.@runat\_07052003\_113851.7658 -NCP-6 -ICPU-3  
-NO\_XLPHY -NO\_MAP -LARGEQUERY -NEG\_SCORES-0 -WAIT -DSPLOCK-100 -LONGLOG  
-DEV\_TIMEOUT-120 -WARN\_TIMEOUT-30 -THREDS-1 -XGAPOP-10 -XGAPEXT-0.5 -FGAPOP-6  
-FGAPEXT-7 -YGAPOP-10 -YGAPEXT-0.5 -DELOP-6 -DELEXT-7

Database : N.Geneseq-101002.\*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	56	12.7	42	24	AAD27031	Human Secreted epi
2	49	11.1	50	8	AAH70336	3'-5' sequence of
3	47	10.6	51	22	AAI78209	Human silent SNP c
4	47	10.6	60	24	ABN39530	Human spliced tran
5	46	10.4	60	24	ABN45376	Human spliced tran
6	46	10.4	65	24	ABN52611	Mouse spliced tran
7	46	10.4	66	24	ABN73405	Bovine embryonic g
8	46	10.4	68	20	ABN33998	Polynucleotide ass
9	46	10.4	71	21	AAA49245	HCY core protein p
10	46	10.4	72	14	AAO51446	Plasmodium gHCV/delt
11	46	10.4	79	14	AAO36160	Human PAC Intron C
12	46	10.4	72	14	AAO36160	Mutagenic primer C
13	45	10.2	48	22	AAI76614	Human silent SNP c
14	45	10.1	65	24	ABN31870	Rat spliced trans
15	44	10.0	50	15	AAO54486	Excitatory amino a
16	44	10.0	50	15	AAO54487	Excitatory amino a
17	44	10.0	60	21	AAA50222	Hybrid leptin/tiss
18	44	10.0	63	21	AAA50221	Hybrid leptin/tiss
19	44	10.0	75	22	AAD09429	Vlons-tpa vaccine
20	43.5	9.8	65	24	ABN31369	Rat spliced trans
21	43.5	9.8	76	21	AAA99033	Human TGC839 PCR p
22	43.5	9.8	80	24	ABL59337	PCR primer for hum
23	43	9.7	50	22	AAI30844	Human SNP oligonuc
24	43	9.7	51	22	AAI75717	Human silent SNP c
25	43	9.7	60	24	ABN38069	Human spliced tran
26	43	9.7	60	24	ABN40440	Human spliced tran
27	43	9.7	65	24	ABN31099	Rat spliced trans
28	43	9.7	66	21	AAA49049	DNA encoding trans
29	43	9.7	66	21	AAA50218	Modified human tis
30	43	9.7	66	22	AAC66513	Modified human tis
31	43	9.7	69	21	AAA50217	Tissue plasminogen
32	43	9.7	69	21	AAA50219	Native human tissu
33	43	9.7	69	21	AAA50220	Modified human tis
34	43	9.7	69	21	AAA50221	Modified human tis
35	43	9.7	69	21	AAA50222	Modified human tis
36	43	9.7	69	21	AAA50223	Modified human tis
37	43	9.7	70	23	ABL58043	Cyclinoxenase-2,
38	43	9.7	75	24	ABL39944	Signal peptide tpa
39	43	9.7	76	21	AAA50338	Synthetic oligonuc
40	43	9.7	78	17	AAI15476	Human tissue-speci
41	43	9.7	78	18	AAI85512	Human tissue-speci
42	43	9.7	78	19	AAV21738	Sense oligomer I u
43	43	9.7	78	22	AAO92203	Antisense oligomer
44	43	9.7	78	22	AAO92204	Synthetic sense ol
45	43	9.7	78	22	AAO92205	Synthetic antisens

## ALIGNMENTS

RESULT 1  
AAD27031  
ID AAD27031 standard; DNA; 42 BP.  
AC AAD27031;  
XX  
DE Human Secreted epithelial colon stromal-1 DNA amplifying PCR primer #5.  
DT 09-APR-2002 (first entry)  
XX  
KW Secreted epithelial colon stromal-1; Seps-1; gene therapy; osteoporosis;  
KW haematopoietic disorder; osteoporosis; osteogenesis imperfecta; cachexia;  
KW Paget's disease; periodontal disease; hypercalcaemia; glomerulonephritis;  
KW diabetes; obesity; osteopathic; cytostatic; nephrotrophic; antidiabetic;  
KW anorectic; immunomodulator; antiproliferative; vulnary; antineoplastic;  
KW gynaecological; anticancer; antineoplastic; cancer; cell therapy;  
KW human; PCR primer; ss.  
XX

```

OS Homo sapiens.
XX
XX WO200198497-A1.
XX
XX 27-DEC-2001.
XX
XX 28-NOV-2000; 2000MO-US32479.
XX
XX 21-JUN-2000; 2000US-0599087.
XX
XX 28-NOV-2000; 2000US-0724000.
XX
XX (AMGE-) AMGEN INC.
XX
XX Polverino AJ, Luethy R;
XX
XX WPI; 2002-122281/16.
XX
XX Secreted epithelial colon stromal-1 polypeptides and nucleic acids,
XX useful for diagnosing, treating and preventing hematopoietic disorder,
XX osteoporosis, Paget's disease, cancer, diabetes -
XX
XX Example 4; Page 89; 134pp; English.
XX
XX The present invention relates to an isolated murine or human secreted
XX epithelial colon stromal-1 (Secs-1) polypeptide, its allelic or splice
XX variant, orthologue, fragment or mutant. Secs-1 gene is used in gene
XX therapy and cell therapy. Secs-1 is useful for identifying a compound
XX which binds to a Secs-1 polypeptide. Secs-1 is useful for treating,
XX preventing or ameliorating a disease condition such as haematopoietic
XX disorder, osteoporosis, osteopetrosis, osteogenesis imperfecta, Paget's
XX disease, petiondontal disease, hypercalcaemia, acute glomerulonephritis,
XX chronic glomerulonephritis, cancer, diabetes, obesity or cachexia. Secs-1
XX is also useful for diagnosing a pathological condition which involves
XX determining the presence or amount of Secs-1 or polypeptide encoded by
XX Secs-1 DNA in a sample; and diagnosis of a pathological condition, or
XX susceptibility to pathological condition based on the presence or amount
XX of expression of the polypeptide. The present sequence is human Secs-1
XX DNA amplifying PCR primer.
XX
XX Sequence 42 BP; 15 A; 10 C; 9 G; 8 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 224 Length: 42
XX Score: 56.00 Matches: 10
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 12.67% Indels: 0
XX DB: 24 Gaps: 0
XX
XX US-09-599-087B-5 (1-81) x AAD27031 (1-42)
XX
XX QY 25 LysARGARProAlaLysAlaTrpSerGly 34
XX |||||||
XX DB 13 AAAGCTGCTCAGCTAAAGCTGCTGAGGC 42
XX
XX RESULT 2
XX AAN70336/c
XX ID AAN70336 standard; DNA: 50 BP.
XX
XX AC AAN70336;
XX
XX XX
XX DT 03-OCT-2002 (updated)
XX DT 02-APR-1991 (first entry)
XX
XX 3'-5' sequence of probe A11 used to detect hepatitis B virus (HBV)
XX DE analyte.
XX
XX XX
XX KM Hepatitis B virus assay; diagnosis; ss.
XX
XX XX
XX OS Synthetic.
XX
XX Key Location/Qualifiers
XX FH misc_feature 1..20
XX FT

```

```

FT FT /*tag= a
FT FT /note="Fluorescein label conjugate binding site"
FT FT misc_feature 21..50
FT FT /*tag= b
FT FT /note="probe segment to HBV"
XX
XX EP225807-A.
XX
XX 16-JUN-1987.
XX
XX 10-DEC-1986; 86EP-0309622.
XX
XX 11-DEC-1985; 85US-0807624.
XX
XX 23-DEC-1986; 86US-0943876.
XX
XX (CHIR-) CHIRON CORP.
XX
XX Ureda MS, Warner B, Horn J;
XX
XX WPI; 1987-165004/24.
XX
XX Detecting specific nucleic acid sequences - by incubating sample
XX with labelling reagent set and capturing reagent set.
XX
XX Example; page 17; 31pp; English.
XX
XX In the example a BglII fragment of hepatitis B virus (AAN70325) was
XX detected. 12 sequences, complementary to different sequences
XX present in HBV were constructed (see AAN70326-n70337). Six of the HBV
XX complementary sequences are joined to a common sequence (A) for
XX complexing with the label conjugate (A'). The other six HBV
XX complementary sequences are joined to a common sequence (B) for
XX complexing with a biotinylated sequence (B') or a third DNA sequence
XX (B'C) for binding to a support (see AAN70338, AAN70339, AAN70340).
XX (Updated on 03-OCT-2002 to add missing OS field.)
XX
XX Sequence 50 BP; 13 A; 8 C; 14 G; 15 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 1.56e+03 Length: 50
XX Score: 49.00 Matches: 8
XX Percent Similarity: 78.57% Conservative: 3
XX Best Local Similarity: 57.14% Mismatches: 3
XX Query Match: 11.09% Indels: 0
XX DB: 8 Gaps: 0
XX
XX US-09-599-087B-5 (1-81) x AAN70336 (1-50)
XX
XX QY 40 GycCySHisArgValProSerProAsnSerThrAsnLeuLys 53
XX |||||||
XX DB 45 TGTTGTCATCAAGGCCCTTCACACACTTGACCAACTTGCA 4
XX
XX RESULT 3
XX AAT78209
XX ID AAT78209 standard; DNA: 51 BP.
XX
XX AC AAT78209;
XX
XX XX
XX DT 09-NOV-2001 (first entry)
XX
XX Human silent SNP containing nucleic acid SEQ:5150.
XX
XX XX
XX KM Human; single nucleotide polymorphism; SNP; genome; gene therapy;
XX KM protein therapy; vaccine; probe; diagnostic assay; detection;
XX KM quantitation; restorative therapy; polymorphic; ds.
XX
XX XX
XX OS Homo sapiens.
XX
XX XX
XX PN WO200140521-A2.
XX
XX 07-JUN-2001.
XX
XX 30-NOV-2000; 2000MO-US32758.
XX
XX FT

```



```

XX 30-NOV-1999: 99US-0168138.
PR 29-NOV-2000: 2000US-0726173.
XX
XX (CDRA-) CDRAGEN CORP.
XX
XX Shinkets RA, Leach M;
XX
XX WPI: 2001-356160/37.
XX
XX polymorphic nucleic acid sequences, useful in genetic testing and
XX therapy -
XX
XX Claim 1; Page 2086; 2653pp; English.
XX
XX AA173060 to AA179867 represent isolated human polymorphic polynucleotide
XX sequences (I), which contain single nucleotide polymorphisms (SNPs).
XX AA453114 to AA453329 represent peptides related to human polymorphic
XX polynucleotide sequences. The sequences can be used in gene and protein
XX therapy, and in vaccine production. (I) and the polypeptides encoded by
XX them may be used in the prevention, diagnosis and treatment of diseases
XX associated with inappropriate expression of polymorphic polypeptides.
XX For example, (I) may be used to treat disorders by rectifying mutations
XX or deletions in a patient's genome that affect the activity of
XX polypeptides by expressing inactive proteins or to supplement the
XX patient's own production of polypeptide. Additionally, (I) and its
XX complementary sequences may also be used as DNA probes in diagnostic
XX assays to detect and quantitate the presence of similar nucleic acids
XX in samples, and therefore which patients may be in need of restorative
XX therapy. The polypeptides encoded by (I) may be used as antigens in the
XX production of antibodies specific for polymorphic polypeptides. The
XX antibodies may also be used to down regulate expression and activity.
XX The antibodies may also be used as diagnostic agents for detecting the
XX presence of polymorphic polypeptides in samples.
XX
XX
XX Sequence 51 BP; 14 A; 13 C; 13 G; 11 T; 0 other:
XX
XX
XX Alignment Scores:
XX Pred. No.: 2.63e+03 Length: 51
XX Score: 47.00 Matches: 8
XX Percent Similarity: 72.73% Conservative: 0
XX Best Local Similarity: 72.73% Mismatches: 3
XX Query Match: 10.63% Indels: 0
XX Db: 22 Gaps: 0
XX
XX US-09-599-087B-5 (1-81) x AA178209 (1-51)
XX
XX Oy 39 LeucycysHisArgValProSerProAsnSer 49
XX ||||| ||| ||||| |||||
XX Db 19 TTGTCCTTGAAACGACGCCCTCAGAGATATCC 51
XX
XX RESULT 4
XX ID ABN39530/C
XX ABN39530 standard; DNA; 60 BP.
XX
XX AC ABN39530;
XX
XX 15-JUL-2002 (first entry)
XX
XX DE Human spliced transcript detection oligonucleotide SEQ ID NO:12278.
XX
XX Human; mouse; rat; splice transcript; detection; RNA transcript;
XX splice variant; transcriptome; oligonucleotide library; ss.
XX
XX OS Homo sapiens.
XX
XX WO200210449-A2.
XX
XX 07-FEB-2002.
XX
XX 20-JUL-2001; 2001WO-IB01903.
XX
XX 28-JUL-2000; 2000US-221607P.

```

```

XX 02-MAY-2001: 2001US-287724P.
XX
XX (COMP-) COMPUGEN INC.
XX
XX Shoshan A, Masserman A, Mintz E, Mintz L, Faigler S;
XX
XX WPI; 2002-257383/30.
XX
XX New oligonucleotide libraries comprising oligonucleotides which
XX selectively hybridize to mRNAs transcribed from a transcription unit of
XX a genome, useful for detecting tissue-, pathology-, and
XX developmental-specific genes
XX
XX Example 1: SEQ ID 12278; 47pp; English.
XX
XX The present invention describes oligonucleotide libraries for detecting
XX messenger RNAs that populate a (sub-)transcriptome, where the
XX (sub-)transcriptome comprises messenger RNAs transcribed from multiple
XX transcription units that populate a genome. The library comprises
XX several oligonucleotides, each capable of hybridizing selectively to a
XX set of messenger RNAs transcribed from a given transcription unit of
XX the genome, which encodes one or more messenger RNA splice variants.
XX The oligonucleotide libraries are useful for detecting mRNAs from a
XX biological sample, in expression profiling studies, in qualitatively or
XX quantitatively characterizing the corresponding transcriptome, and in
XX detecting RNA transcripts and splice variants of human or animal
XX libraries. The libraries may also be used as specialized mini
XX libraries to detect transcripts of a sub-transcriptome under a
XX particular biological or pathological state, and so allowing the
XX detection of tissue- and pathology-specific genes such as those genes
XX only expressed in specific tissue under a specific pathological
XX condition; to detect developmental specific genes; and to detect RNA
XX transcripts and splice variants of a transcriptome of a patient suffering
XX from a particular disorder. ABN27253 to ABN59589 represent
XX oligonucleotide sequences from rats, humans and mice, which are used in
XX the exemplification of the present invention.
XX N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 60 BP; 13 A; 23 C; 9 G; 15 T; 0 other;
XX
XX Alignment Scores:
XX Pred. NO.: 3.23e+03 Length: 60
XX Score: 47.00 Matches: 9
XX Percent Similarity: 71.43% Conservative: 1
XX Best Local Similarity: 64.29% Mismatches: 4
XX Query Match: 10.63% Indels: 0
XX DB: 24 Gaps: 0
XX
XX US-09-599-087B-5 (1-81) x ABN393530 (1-60)
XX
XX 21 serthrglmglylsyargargproalalyalatrpsrgly 34
XX |||:::||||| ||||||| |||||||
XX 56 TCGTCTGAAGGCGCTTAGGAGACCTTGGCTTGGATGGAGTGGG 15
XX
XX RESULT 5
XX ID ABN45376/c ABN45376/c
XX AC ABN45376;
XX
XX 15-JUL-2002 (first entry)
XX
XX Human spliced transcript detection oligonucleotide seq ID NO:18124.
XX
XX Human; mouse; rat; splice transcript; detection; RNA transcript;
XX splice variant; transcriptome; oligonucleotide library; ss.
XX
XX Homo sapiens.
XX
XX WO200210449-A2.
XX

```

PD 07-FEB-2002.  
XX  
XX 20-JUL-2001; 2001WO-IB01903.  
XX  
XX 28-JUL-2000; 2000US-221607P.  
PR 02-MAY-2001; 2001US-287724P.  
XX  
XX (COMP-) COMPUGEN INC.  
XX  
XX Shoshan A, Masserman A, Mintz E, Mintz L, Faigler S;  
XX WPI; 2002-257383/30.  
XX  
XX New oligonucleotide libraries comprising oligonucleotides which  
PT selectively hybridize to mRNAs transcribed from a transcription unit of  
PT a genome, useful for detecting tissue-, pathology-, and  
PT developmental-specific genes -  
XX  
XX Example 1; SEQ ID 18124; 47pp; English.  
XX  
XX The present invention describes oligonucleotide libraries for detecting  
CC messenger RNAs that populate a (sub-)transcriptome, where the  
CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple  
CC transcription units that populate a genome. The library comprises  
CC several oligonucleotides, each capable of hybridizing selectively to a  
CC set of messenger RNAs transcribed from a given transcription unit of  
CC the genome, which encodes one or more messenger RNA splice variants.  
CC The oligonucleotide libraries are useful for detecting mRNAs from a  
CC biological sample, in expression profiling studies, in qualitatively or  
CC quantitatively characterizing the corresponding transcriptome, and in  
CC detecting RNA transcripts and splice variants of human or animal  
CC transcriptomes. The libraries may also be used as specialised mini  
CC libraries to detect transcripts of a sub-transcriptome under a  
CC particular biological or pathological state, and so allowing the  
CC detection of tissue- and pathology-specific genes such as those genes  
CC only expressed in specific tissue under a specific pathological  
CC condition; to detect developmental specific genes; and to detect RNA  
CC transcripts and splice variants of a transcriptome of a patient suffering  
CC from a particular disorder. ABN27253 to ABN59589 represent  
CC oligonucleotide sequences from rats, humans and mice, which are used in  
CC the exemplification of the present invention.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 60 BP; 14 A; 16 C; 21 G; 9 T; 0 other;  
SQ  
Alignment Scores:  
Pred. No.: 4.13e+03 Length: 60  
Score: 46.00 Matches: 7  
Percent Similarity: 63.64% Conservative: 0  
Best Local Similarity: 63.64% Mismatches: 4  
Query Match: 10.41% Indels: 0  
DB: 24 Gaps: 0  
US-09-599-087B-5 (1-81) x ABN45376 (1-60)  
QY 62 ProCysLysLeuGluProGluProArgLeuTyr 72  
DB 55 CCGTGCAGTCTCGTTCGAGCCGAGATCTCTGG 23  
RESULT 6  
ABN52611/c  
ID ABN52611 standard; DNA; 65 BP.  
XX  
XX ABN52611:  
XX  
XX 15-JUL-2002 (first entry)  
XX  
XX Mouse spliced transcript detection oligonucleotide SEQ ID NO:25359.  
XX  
XX Human: mouse; rat; splice transcript; detection; RNA transcript;  
KM splice variant; transcriptome; oligonucleotide library; ss.

XX  
XX Mus musculus.  
XX  
XX WO200210449-A2.  
XX  
XX  
XX 07-FEB-2002.  
XX  
XX 20-JUL-2001; 2001WO-IB01903.  
XX  
XX 28-JUL-2000; 2000US-221607P.  
PR 02-MAY-2001; 2001US-287724P.  
XX  
XX (COMP-) COMPUGEN INC.  
XX  
XX Shoshan A, Masserman A, Mintz E, Mintz L, Faigler S;  
XX WPI; 2002-257383/30.  
XX  
XX New oligonucleotide libraries comprising oligonucleotides which  
PT selectively hybridize to mRNAs transcribed from a transcription unit of  
PT a genome, useful for detecting tissue-, pathology-, and  
PT developmental-specific genes -  
XX  
XX Example 1; SEQ ID 25359; 47pp; English.  
XX  
XX The present invention describes oligonucleotide libraries for detecting  
CC messenger RNAs that populate a (sub-)transcriptome, where the  
CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple  
CC transcription units that populate a genome. The library comprises  
CC several oligonucleotides, each capable of hybridizing selectively to a  
CC set of messenger RNAs transcribed from a given transcription unit of  
CC the genome, which encodes one or more messenger RNA splice variants.  
CC The oligonucleotide libraries are useful for detecting mRNAs from a  
CC biological sample, in expression profiling studies, in qualitatively or  
CC quantitatively characterizing the corresponding transcriptome, and in  
CC detecting RNA transcripts and splice variants of human or animal  
CC transcriptomes. The libraries may also be used as specialised mini  
CC libraries to detect transcripts of a sub-transcriptome under a  
CC particular biological or pathological state, and so allowing the  
CC detection of tissue- and pathology-specific genes such as those genes  
CC only expressed in specific tissue under a specific pathological  
CC condition; to detect developmental specific genes; and to detect RNA  
CC transcripts and splice variants of a transcriptome of a patient suffering  
CC from a particular disorder. ABN27253 to ABN59589 represent  
CC oligonucleotide sequences from rats, humans and mice, which are used in  
CC the exemplification of the present invention.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 65 BP; 20 A; 16 C; 13 G; 16 T; 0 other;  
SQ  
Alignment Scores:  
Pred. No.: 4.58e+03 Length: 65  
Score: 46.00 Matches: 8  
Percent Similarity: 78.57% Conservative: 3  
Best Local Similarity: 57.14% Mismatches: 3  
Query Match: 10.41% Indels: 0  
DB: 24 Gaps: 0  
US-09-599-087B-5 (1-81) x ABN52611 (1-65)  
QY 38 ArgLeuCysCysHisArgValProSerProAsnSerThrAsn 51  
DB 61 AGAGTTTGACACTATAGAGATCCCAAGTGGTTCATATAAAT 20  
RESULT 7  
ABN73405  
ID ABN73405 standard; cDNA; 66 BP.  
XX  
XX ABN73405;  
XX  
XX 03-JUL-2002 (first entry)



XX (MEDT-) MEDICAL RES COUNCIL.  
PA Hope G, McLauchlan J;  
XX WPI; 2000-400040/34.  
XX  
PT Protein comprising a lipid globule targeting sequence consisting of a  
PT hepatitis C virus core protein, useful for targeting a protein of  
PT interest to lipid globules which are subsequently secreted into animal  
XX milk -  
PS Example; Page 32; 60pp; English.  
XX  
CC The present sequence is a mutagenic primer used in the construction of a  
CC plasmid containing a mutated version of the hepatitis C virus core  
CC protein gene. The core protein has been shown to be localised to lipid  
CC droplets within cells. This means that fragments of the protein, in  
CC particular those comprising amino acids 125-144 or 161-166, can be used  
CC to target proteins to lipid droplets in milk. This provides a simple and  
CC efficient means of obtaining proteins of interest, using transgenic  
CC animals. Proteins of interest include those implicated in disease and  
CC antigenic polypeptides for use as vaccines.  
XX  
SQ Sequence 71 BP; 13 A; 20 C; 22 G; 16 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 5.13e+03 Length: 71  
Score: 46.00 Matches: 9  
Percent Similarity: 68.75% Conservative: 2  
Best Local Similarity: 56.25% Mismatches: 5  
Query Match: 10.41% Indels: 0  
DB: 21 Gaps: 0  
  
US-09-599-087B-5 (1-81) x AAA46147 (1-71)  
OY 25 LysArgArgProAlaLysAlaTrpSerGlyArgArgThrArgLysCys 40  
DB 19 CGTGGGCGCCCTCTTGTCCGGTTCTGGAAGACGGTGAACATATCC 66  
  
RESULT 10  
ID AAA46147 standard; cDNA; 71 BP.  
AC AAA46147;  
XX  
XX 27-OCT-2000 (first entry)  
XX  
DE Plasmid pghcv/delta145-154 insertion oligonucleotide.  
XX  
XX Hepatitis C virus; core protein; HCV; lipid globule; chronic hepatitis;  
XX liver disease; ADRP displacement; insertion oligonucleotide; ss.  
OS Synthetic.  
XX  
XX WO2000031532-A1.  
XX  
XX 02-JUN-2000.  
XX  
XX 23-NOV-1999; 99WO-GB03906.  
XX  
XX 26-NOV-1998; 98GB-0025951.  
XX  
XX (MEDT-) MEDICAL RES COUNCIL.  
XX  
XX Hope G, McLauchlan J;  
XX  
XX WPI; 2000-400181/34.  
XX  
XX Identifying a substance for treating or preventing a viral infection  
XX such as hepatitis C virus, comprises determining if the substance  
XX disrupts a lipid globule target sequence to lipid globule interaction

XX Examples; Page 37; 74pp; English.  
XX  
XX The present sequence is an oligonucleotide which was inserted into  
XX a plasmid which contained part of hepatitis C virus core protein gene.  
XX The virus causes chronic hepatitis and liver disease in humans. It is  
XX thought that it does this by associating with intracellular lipid  
XX droplets and downregulating the expression of adipocyte-specific  
XX differentiation-related protein (ADRP), which is probably required for  
XX the maintenance of lipid droplets. The core protein can be used to  
XX identify substances capable of disrupting this interaction, which are  
XX thus capable of treating or preventing infection by the hepatitis C virus  
XX and its accompanying effects. This oligonucleotide was used in the  
XX construction of mutants of the core protein, which were used to determine  
XX their efficiency of infection compared to the wild-type.  
XX  
SQ Sequence 71 BP; 13 A; 20 C; 22 G; 16 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 5.13e+03 Length: 71  
Score: 46.00 Matches: 9  
Percent Similarity: 68.75% Conservative: 2  
Best Local Similarity: 56.25% Mismatches: 5  
Query Match: 10.41% Indels: 0  
DB: 21 Gaps: 0  
  
US-09-599-087B-5 (1-81) x AAA46147 (1-71)  
OY 25 LysArgArgProAlaLysAlaTrpSerGlyArgArgThrArgLysCys 40  
DB 19 CGTGGGCGCCCTCTTGTCCGGTTCTGGAAGACGGTGAACATATCC 66  
  
RESULT 11  
ID AAQ51446/c  
XX  
XX AAQ51446 standard; DNA; 72 BP.  
AC AAQ51446;  
XX  
XX 20-MAY-1994 (first entry)  
XX  
XX Human FACC Intron 9 3' region.  
XX  
XX  
XX Fanconi Anemia Group C; FACC; complementing cDNA; variant; diagnosis;  
XX Open Reading Frame; Fanconi anemia; gene therapy; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO9322435-A.  
XX  
XX 11-NOV-1993.  
XX  
XX 27-APR-1993; 93WO-CA00178.  
XX  
XX 29-APR-1992; 92US-0876285.  
XX  
XX 21-OUL-1992; 92US-081813.  
XX  
XX 15-JAN-1993; 93US-0003963.  
XX  
XX (HOSP-) HOSPITAL FOR SICK CHILDREN.  
XX  
XX (UNME-) UNITED MEDICAL & DENTAL SCHOOL GUYS.  
XX  
XX Buchwald M, Mathew CG, Strathdee CA, Wevrick R;  
XX  
XX WPI; 1993-368794/46.  
XX  
XX Human cDNA which complements Fanconi Anaemia gp. C - used to  
XX develop prods. for use in diagnosis, study and therapy of Fanconi  
XX Anaemia  
XX  
XX Claim 1; Page 111; 137pp; English.  
XX  
XX The sequences given in AAQ51429-54 represent the 5' and 3' splice  
XX regions of the introns from the Fanconi Anemia Group C Complementing  
XX (FACC) DNA. Three cDNA molecules which are cellular variants of a

score: 40.00  
Percent Similarity: 52.948

Best Local Similarity: 61.54%

Query Match: 10.18% Indels: 0  
DB: 22 Gaps: 0

US-09-599-087b-5 (1-81) x AA176614 (1-48)

OY 67 ProGluProArgLeuTrpValValProGlyAlaLeuPro 79  
||| |||||  
:::|||||

DB 44 CCCAGTCCCGACCTCTGCTCCAGGCGCATCCCG 6

RESULT 14  
ABN31870  
ID ABN31870 standard; DNA; 65 BP.  
XX  
AC ABN31870;  
XX  
XX 15-JUL-2002 (first entry)  
XX  
XX Rat spliced transcript detection oligonucleotide SEQ ID NO:4618.  
XX  
XX Human; mouse; rat; splice transcript; detection; RNA transcript;  
XX  
XX splice variant; transcriptome; oligonucleotide library; ss.  
XX  
XX Rattus norvegicus.  
XX  
XX WO200210449-A2.  
XX  
XX 07-FEB-2002.  
XX  
XX 20-JUL-2001; 2001WO-IB01903.  
XX  
XX 28-JUL-2000; 2000US-221607P.  
XX  
XX 02-MAY-2001; 2001US-287724P.  
XX  
XX (COMP-) COMPUGEN INC.  
XX  
XX Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;  
XX  
XX WPI; 2002-257383/30.  
XX  
XX New oligonucleotide libraries comprising oligonucleotides which  
XX  
XX selectively hybridize to mRNAs transcribed from a transcription unit of  
XX  
XX a genome, useful for detecting tissue-, pathology-, and  
XX  
XX developmental-specific genes

Example 1; SEQ ID 4618; 47pp; English.

The present invention describes oligonucleotide libraries for detecting messenger RNAs that populate a (sub-)transcriptome, where the (sub-)transcriptome comprises messenger RNAs transcribed from multiple transcription units that populate a genome. The library comprises several oligonucleotides, each capable of hybridizing selectively to a set of messenger RNAs transcribed from a given transcription unit of the genome, which encodes one or more messenger RNA splice variants. The oligonucleotide libraries are useful for detecting mRNAs from a biological sample, in expression profiling studies, in qualitatively or quantitatively characterizing the corresponding transcriptome, and in detecting RNA transcripts and splice variants of human or animal transcriptomes. The libraries may also be used as specialized mini libraries to detect transcripts of a sub-transcriptome under a particular biological or pathological state, and so allowing the detection of tissue- and pathology-specific genes such as those genes only expressed in specific tissue under a specific pathological condition; to detect developmental specific genes; and to detect RNA transcripts and splice variants of a transcriptome of a patient suffering from a particular disorder. ABN27253 to ABN59589 represent oligonucleotide sequences from rats, humans and mice, which are used in the exemplification of the present invention.

N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 65 BP; 19 A; 20 C; 13 G; 13 T; 0 other;

Alignment Scores: 6.63e+03 Length: 65  
Pred. No.: 44.50 Matches: 9  
Score: 70.59% Conservative: 3  
Percent Similarity: 52.94% Mismatches: 2  
Best Local Similarity: 10.07% Indels: 3  
Query Match: 24 Gaps: 1

US-09-599-087b-5 (1-81) x ABN31870 (1-65)

OY 43 ArgValProSerProAsnser-----ThrAsnLeuLysGlyHisHis 56  
|||:|||||  
1 AGAATTCCAGCCCTGCTTCAGAGTGCACACCTTCAGAGTCATAT 51

DB 1 AGAATTCCAGCCCTGCTTCAGAGTGCACACCTTCAGAGTCATAT 51

RESULT 15  
AA054486/C  
ID AA054486 standard; cDNA; 50 BP.  
XX  
AC AA054486;  
XX  
XX 07-JUL-1994 (first entry)  
XX  
XX Excitatory amino acid receptor (EAA4a) coding sequence fragment.  
XX  
XX Excitatory amino acid receptor; EAA4a; screening; detection;  
XX  
XX identification; CNS; nervous disorders; antibody; oocyte; membrane  
XX  
XX protein; ion-channel; therapeutics; ss.  
XX  
XX Homo sapiens.  
XX  
XX EP578409-A.  
XX  
XX 12-JAN-1994.  
XX  
XX 24-JUN-1993; 93EP-0304961.  
XX  
XX 24-JUN-1992; 92US-0903456.  
XX  
XX (ELLI/) ELLIOTT C E.  
XX  
XX (KAMBOJ/) KAMBOJ R.  
XX  
XX (NNTT/) NNTT S L.  
XX  
XX Kamboj R, Nutt SL, Elliott CE;  
XX  
XX WPI; 1994-010248/02.  
XX  
XX Excitatory amino acid for receptors with kainate binding activity  
XX  
XX - used to test ligands for CNS receptor interaction to identify  
XX  
XX cpds. useful against CNS disorders

Disclosure; Figure 3b; 37pp; English.

The excitatory amino acid receptor (See AA05367) can be used to screen ligands. The ligands can be assayed for interaction by incubating them with cells, preferably oocytes, expressing the receptor or with membrane preparations from these cells, and assessing any interaction by determining receptor-ligand binding or ligand-mediated ion channel activation. This method is useful in the identification of therapeutics useful to treat CNS disorders in humans. Fragments of the receptor are useful for structural investigations and to raise antibodies. This fragment of the coding sequence (AA054482) corresponds to bases 1971 to 2020.

Sequence 50 BP; 11 A; 11 C; 15 G; 13 T; 0 other;

Alignment Scores: 5.35e+03 Length: 50  
Pred. No.: 44.00 Matches: 7  
Score: 60.00% Conservative: 2  
Percent Similarity: 46.67% Mismatches: 6  
Best Local Similarity: 9.95% Indels: 0  
Query Match: 15 Gaps: 0

Fri May 9 16:38:51 2003

us-09-599-087b-5.rng

Page 9

US-09-599-087b-5 (1-81) x AAQ54486 (1-50)

Qy 34 GlyArgGlyThrArgLeuCyCysHisArgValProSerProAsn 48  
|||::|  
Db 46 GGCATGAGCTCAGAACCTTGTCTGATGAGAGCTCCAACTCCAAAC 2

Search completed: May 8, 2003, 22:43:14  
Job time : 207 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 8, 2003, 22:36:33 ; Search time 1229 Seconds  
(without alignments)

1918.086 Million cell updates/sec

Title: US-09-599-087B-5

Perfect score: 442  
Sequence: 1 NRLVLSLLCILLICFSIF.....PCKLPERRIMVPGALPOV 81

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 463318

Minimum DB seq length: 25  
Maximum DB seq length: 80

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL-frame+pn.model -DEV-xlp  
-O/cgn2.1/USPTO.spool/US09599087/runat.07052003.113852.7668/app.query.fasta.1.263  
-DB-GenEmbl -QFMT-fastap -SUFFIX-rige -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITs-bits -START=1 -END=1 -MATRIX-biosum62 -TRANS-human40.cel -LIST=45  
-DOCALIGN=200 -THR SCORE-pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE-LOCAL  
-OUTFMT-pco -NORM-ext -HEAPSIZE=500 -MINLEN=25 -MAXLEN=80  
-USER-US09599087@cgn.1.1.1319 -runat.07052003.113852.7668 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_XMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_pro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_ju:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_pro:\*  
27: em\_sts:\*  
28: em\_un:\*

29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_in:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pin:\*  
35: em\_htg\_rod:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	59.5	13.5	76	6 A07238	A07238 Probe for v
2	56	12.7	42	6 AX342232	AX342232 Sequence
3	53	12.0	79	6 I86930	I86930 Sequence 19
4	51.5	11.7	69	9 HS090998	HS0998 Homo sapien
5	49	11.1	66	10 MUSTCAY	M36677 Mouse T-cell
6	49	11.1	70	10 RNTCRAJ39	Y09178 R.norvegicus
7	48	10.9	76	9 S7538518	S75420 collagen ty
8	47	10.6	51	6 AX161822	AX161822 Sequence
9	46.5	10.5	48	6 AR077574	AR077574 Sequence
10	46	10.4	71	6 AX025072	AX025072 Sequence
11	46	10.4	71	6 AX025094	AX025094 Sequence
12	46	10.4	72	6 I71481	I71481 Sequence 22
13	46	10.4	78	9 HSIGKAB01	X54166 Human BL ty
14	45.5	10.3	69	6 AR005307	AR005307 Sequence
15	45	10.2	39	6 AR077518	AR077518 Sequence
16	45	10.2	48	6 AX160227	AX160227 Sequence
17	45	10.2	76	6 AF118405	AF118405 Homo sapl
18	44	10.0	42	6 AR014091	AR014091 Sequence 7
19	44	10.0	50	6 I28909	I28909 Sequence 7
20	44	10.0	50	6 I28910	I28910 Sequence 8
21	44	10.0	51	6 AR077572	AR077572 Sequence
22	44	10.0	51	6 AR077575	AR077575 Sequence
23	44	10.0	66	6 A28858	A28858 Oligonucleo
24	44	10.0	79	5 DUKAETC04	M55135 Peking duck
25	43	9.7	46	6 A06605	A06605 Oligonucleo
26	43	9.7	46	6 A10221	A10221 adapter. 1/
27	43	9.7	51	6 AX159330	AX159330 Sequence
28	43	9.7	66	6 AX040385	AX040385 Sequence
29	43	9.7	72	9 HSADD2531	AF000111 Homo sapl
30	43	9.7	76	1 MCTRFM	X16759 Mycoplasma
31	43	9.7	76	1 MCTRFM	K00312 Mycoplasma
32	43	9.7	76	6 A07239	A07239 Probe for t
33	43	9.7	78	6 AR000573	AR000573 Sequence
34	43	9.7	78	6 AR000574	AR000574 Sequence
35	43	9.7	78	6 AR208886	AR208886 Sequence
36	43	9.7	78	6 AR208887	AR208887 Sequence
37	43	9.7	78	6 BD003625	BD003625 A polynuc
38	43	9.7	78	6 BD003626	BD003626 A polynuc
39	43	9.7	79	6 AR050027	AR050027 Sequence
40	42.5	9.6	69	9 HS091209	U91209 Homo sapien
41	42	9.5	30	6 AX003724	AX003724 Sequence
42	42	9.5	47	6 AX003724	AX003724 Sequence
43	42	9.5	51	6 AX159329	AX159329 Sequence
44	42	9.5	51	6 AX165793	AX165793 Sequence
45	42	9.5	52	9 HS19	X68693 H.sapiens m

RESULT 1

## ALIGNMENTS



A07238  
LOCUS A07238 76 bp DNA linear PAT 24-AUG-1993  
DEFINITION Probe for vPA-beta.  
ACCESSION A07238  
VERSION A07238.1 GI:413006  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1 (bases 1 to 76)  
AUTHORS Baldus, B., Donner, P., Schleuning, W.D., Alagon, A., Boisdol, W., Kraetzschmar, J.R., Haendler, B.J. and Langer, G.  
TITLE Novel thrombolytic  
JOURNAL Patent: EP 0383417-A 1 22-AUG-1990;  
FEATURES  
source 1. 76  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
BASE COUNT 14 a 17 c 24 g 21 t  
ORIGIN

Alignment Scores:  
Pred. No.: 566 Length: 76  
Score: 59.50 Matches: 14  
Percent Similarity: 66.67% Conservative: 4  
Best Local Similarity: 51.85% Mismatches: 7  
Query Match: 13.46% Indels: 2  
DB: 6 Gaps: 1

US-09-599-087b-5 (1-81) x A07238 (1-76)

QY 9 LeuLeuGysileLeuLeuLeuGysPheSerIlePheSer-ThrGluGlyPysArgArgPr 28  
Db 1 CTGTTGTGTACTGCTTTGTGGAGACAGTCTTCGTGCCAGGC--AGGAAC 57  
QY 28 CAlAlYsAlATrPserGly 34  
Db 58 TACAGGCATTGGCAGGG 76

RESULT 2  
AX342232 42 bp DNA linear PAT 12-JAN-2002  
LOCUS AX342232  
DEFINITION Sequence 18 from Patent WO0198497.  
ACCESSION AX342232  
VERSION AX342232.1 GI:18151770  
KEYWORDS  
SOURCE synthetic construct.  
ORGANISM synthetic construct.  
REFERENCE 1  
AUTHORS Polverino, A.J. and Luethy, R.  
TITLE Secreted epithelial colon stromal-1 polypeptides, nucleic acids encoding the same and uses thereof  
JOURNAL Patent: WO 0198497-A 18 27-DEC-2001;  
FEATURES  
source 1. 42  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="PCR primer corresponding to human SECS-1"  
BASE COUNT 15 a 10 c 9 g 8 t  
ORIGIN

Alignment Scores:  
Pred. No.: 642 Length: 42  
Score: 56.00 Matches: 10  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 12.67% Indels: 0  
DB: 6 Gaps: 0

US-09-599-087b-5 (1-81) x AX342232 (1-42)

QY 25 LysArgPrAlaLysAlaTrpSerGly 34  
Db 13 AAAGTCGTCAGCTAAAGCCTGGTCAGGC 42

RESULT 3  
LOCUS 186930 79 bp DNA linear PAT 10-JUN-1998  
DEFINITION Sequence 19 from patent US 5702931.  
ACCESSION 186930  
VERSION 186930.1 GI:3206648  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 79)  
AUTHORS Andrews, W.H., Moser, M.J. and Villander, L.R.  
TITLE Mutagenesis methods and compositions  
JOURNAL Patent: US 5702931-A 19 30-DEC-1997;  
FEATURES  
source 1. 79  
/organism="unknown"  
BASE COUNT 10 a 28 c 31 g 10 t  
ORIGIN

Alignment Scores:  
Pred. No.: 2.45e+03 Length: 79  
Score: 53.00 Matches: 9  
Percent Similarity: 55.00% Conservative: 2  
Best Local Similarity: 45.00% Mismatches: 9  
Query Match: 11.99% Indels: 0  
DB: 6 Gaps: 0

US-09-599-087b-5 (1-81) x 186930 (1-79)

QY 29 AlAlYsAlATrPserGlyArgArgTrhArgLeuGysCysHisArgValProSerProAsn 48  
Db 15 GCCTCGGGCTGGGGGGGCTCGCCAGAGCCGCTGTGCCACCTCCACCTTGGCCAGT 74

RESULT 4  
HSU90998/c 69 bp mRNA linear PRI 15-JUN-1997  
LOCUS HSU90998  
DEFINITION Homo sapiens clone FE1389 T-cell receptor delta chain (TCRV1J2)  
ACCESSION U90998  
VERSION U90998.1 GI:2239403  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
REFERENCE 1  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLE The TCR-delta repertoire in human intestine undergoes characteristic changes during fetal to adult development  
JOURNAL J. Immunol. 158 (12), 5632-5641 (1997)  
MEDLINE 97334214  
PUBMED 9190911  
REFERENCE 2 (bases 1 to 69)  
AUTHORS Holtmeier, W., Witthoft, T., Hennemann, A., Winter, H.S. and Kagnoff, M.F.  
TITLE Direct Submission  
JOURNAL Submitted (27-FEB-1997) Department of Medicine, University of Frankfurt, Theodor-Stern Kai #7, Frankfurt 60590, Germany  
FEATURES  
source 1. 69  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="14"  
/clone="FE1389"  
/tissue="small intestine"  
/dev\_stage="fetus; 20 weeks gestation"

```

gene      <1..>69
          /gene="TCRDV1J2"
CDS       <1..>69
          /gene="TCRDV1J2"
          /note="rearranged; contains CDR3 domain, 11 amino acids
          (calculation according to: Rock, E., J. Exp. Med.,
          179:323-328, 1994)"
          /codon_start=1
          /product="T-cell receptor delta chain"
          /protein_id="AAC51561.1"
          /db_xref="GI:2239404"
          /translation="LEDSAKYFCALEIGDTVAQLF"
          <1..>69
          /gene="TCRDV1J2"
BASE COUNT 18 a 14 c 19 g 18 t
ORIGIN
Alignment Scores:
Pred. No.: 2.95e+03 Length: 69
Score: 51.50 Matches: 9
Percent Similarity: 73.33% Conservative: 2
Best Local Similarity: 60.00% Mismatches: 3
Query Match: 11.65% Indels: 1
DB: 9 Gaps: 1
US-09-599-087b-5 (1-81) x HS090998 (1-69)
QY 39 Leucocyte-specific protein tyrosine kinase 52
Db 63 TTGTGCTGTACCCATCCCGAGTCCCAAGACAAAGTA 19
RESULT 5
LOCUS MUSTCAYJ 66 bp mRNA linear ROD 27-APR-1993
DEFINITION Mouse T-cell receptor alpha-chain mRNA J-C region, partial cds.
ACCESSION M38677.1 GI:201214
VERSION M38677.1 J-region; T-cell receptor.
KEYWORDS Mouse, cDNA to mRNA.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE 1 (bases 1 to 66)
AUTHORS Yague, J., Blackman, M., Born, W., Marrack, P., Kappler, J. and
Palmer, E.
TITLE The structure of V alpha and J alpha segments in the mouse
JOURNAL Nucleic Acids Res. 16 (23), 11355-11364 (1988)
MEDLINE 89083566
PUBMED 2849763
FEATURES
source Location/Qualifiers
1..66
/organism="Mus musculus"
/db_xref="taxon:10090"
1..66
/partial
/codon_start=1
/product="T-cell receptor alpha chain"
/db_xref="GI:201215"
/translation="TTASLGKLOFGGTGVVVPDI"
BASE COUNT 18 a 16 c 13 g 13 t
ORIGIN
Alignment Scores:
Pred. No.: 4.86e+03 Length: 66
Score: 49.00 Matches: 8
Percent Similarity: 80.00% Conservative: 0
Best Local Similarity: 80.00% Mismatches: 2
Query Match: 11.09% Indels: 0
DB: 10 Gaps: 0
US-09-599-087b-5 (1-81) x MUSTCAYJ (1-66)

```

```

QY 63 CysLysLeuGluProGluProArgLeuTrip 72
Db 23 TGCAGTTTGGAACAGACCAAGGAGTGTGG 52
RESULT 6
LOCUS RNTCRAJ39 70 bp mRNA linear ROD 06-JAN-1998
DEFINITION R. norvegicus TCR alpha joining region, clone-library VA82P39.
VERSION Y09178
KEYWORDS TCR alpha joining region.
SOURCE Rattus norvegicus.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE 1 (bases 1 to 70)
AUTHORS Torres-Nagel, N., Deutschland, A., Herrmann, T., Arden, B. and
Hunig, T.
TITLE Control of TCR V alpha-mediated positive repertoire selection and
alloreactivity by differential J alpha usage and CDR3 alpha
composition
JOURNAL Int. Immunol. 9 (10), 1441-1452 (1997)
MEDLINE 98013626
PUBMED 9352349
REFERENCE 2 (bases 1 to 70)
AUTHORS Torres-Nagel, N.E.
TITLE Submitted (31-OCT-1996) N.E. Torres-Nagel, Institut de Virology and
Immunobiology, Department of Immunology, Versbacherstr. 7, 97078
Wuerzburg, FRG
FEATURES
source Location/Qualifiers
1..70
/organism="Rattus norvegicus"
/strain="LEW.1F"
/db_xref="taxon:10116"
/naplotype="RT1F"
/cell_type="peripheral T cell"
/clone_lib="VA82P39"
/dev_stage="adult"
/rearranged
1..70
/gene="TCRAJ"
/gene="TCRAJ"
<1..>70
/codon_start=1
/protein_id="CAA70377.1"
/db_xref="GI:1785816"
/translation="AVEIATGFGKLOFGGTGVVVP"
<1..>70
/gene="TCRAJ"
/note="CDR3 region and J alpha region"
/evidence="experimental"
BASE COUNT 16 a 13 c 24 g 17 t
ORIGIN
Alignment Scores:
Pred. No.: 5.18e+03 Length: 70
Score: 49.00 Matches: 8
Percent Similarity: 80.00% Conservative: 0
Best Local Similarity: 80.00% Mismatches: 2
Query Match: 11.09% Indels: 0
DB: 10 Gaps: 0
US-09-599-087b-5 (1-81) x RNTCRAJ39 (1-70)
QY 63 CysLysLeuGluProGluProArgLeuTrip 72
Db 32 TGCAGTTTGGAACAGACCAAGGAGTGTGG 61
RESULT 7
LOCUS S75385S18 76 bp DNA linear PRI 02-JUN-2000

```

DEFINITION collagen type VI alpha 1(VI) chain (exons 3-19 and exon/intron junctions) [human, Genomic, 76 nt, segment 18 of 18].

ACCESSION S75420

VERSION S75420.1 GI:241983

KEYWORDS 18 of 18

SEGMENT Homo sapiens

SOURCE Homo sapiens

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 76)

AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Eutheria; Primates; Catarrhini; Homnidae; Homo. Saitta, B., Wang, Y.M., Renkart, L., Zhang, R.Z., Pan, T.C., Timpl, R. and Chu, M.L.

TITLE The exon organization of the triple-helical coding regions of the human alpha 1(VI) and alpha 2(VI) collagen genes is highly similar

JOURNAL Genomics 11 (1), 145-153 (1991)

MEDLINE 92112205

PubMed 1765372

REMARK GenBank staff at the National Library of Medicine created this entry [NCBI g1dbsq 75420] from the original journal article. This sequence comes from Figure 2.

Map location: chromosome 21.

Location/Qualifiers

1..76

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="21"

join(S75385.1:<30..74,S75388.1:31..57,S75390.1:31..57,S75392.1:31..75,S75394.1:31..84,S75396.1:31..93,S75398.1:31..93,S75400.1:31..84,S75402.1:31..66,S75404.1:31..93,S75406.1:31..81,S75408.1:31..93,S75410.1:31..93,S75412.1:31..81,S75414.1:31..66,S75416.1:31..93,S75418.1:31..96,31..>66)

/product="collagen type VI alpha 1(VI) chain"

join(S75385.1:<30..74,S75388.1:31..57,S75390.1:31..57,S75392.1:31..75,S75394.1:31..84,S75396.1:31..93,S75398.1:31..93,S75400.1:31..84,S75402.1:31..66,S75404.1:31..93,S75406.1:31..81,S75408.1:31..93,S75410.1:31..93,S75412.1:31..81,S75414.1:31..66,S75416.1:31..93,S75418.1:31..96,31..>66)

/gene="collagen type VI alpha 1(VI) chain"

/note="This sequence comes from Figure 2; conceptual translation differs from the translation provided in the manuscript"

/codon\_start=1

/product="collagen type VI alpha 1(VI) chain"

/protein\_id="AAB20835.2"

/db\_xref="GI:8176524"

/translation="GRPGDLAGPVGYOGMKGEKSGRGPYKNGEKGRKIDGVGVGEMGYGLPCGKSGPGFGLGCGPPKGPDPGAGLKEKGPADGEAGPARGGPSGDEGPAGPPEKSGEAGDGNPGPAGPGERGPGRGRGPRGPDGSAAGROGGRGPGVGVCPDGEAGPLGPKGYRDEGPGSGAGAGAPRPDPRLMGERGDEGPAKGTGEGFPFGYRGNKGAAGINCKGYPLGKGDEGAGDPDNDIAPRGVKGAKGRPEGPQGPFGHGGPPGP"

31..>66

/number=20

exon

BASE COUNT 15 a 26 c 24 g 11 t

ORIGIN

Alignment Scores:

Pred. No.: 7.04e+03 Length: 76

Score: 48.00 Matches: 9

Percent Similarity: 59.09% Conservative: 4

Best Local Similarity: 40.91% Mismatches: 9

Query Match: 10.86% Indels: 0

DB: 9 Gaps: 0

US-09-599-087b-5 (1-81) x S75385S18 (1-76)

OY 15 LeucysPheSerIlePheSerThrCluGlyLysArgArgProAlaLysAlaTyrSerGly 34

DB 6 CTCCTCTCTCCGTTCTCGACAGGACCCCGAGGACACCAAGACCGCTGGGCCGGA 65

OY 35 ArgArg 36

DB 66 CGTAG 71

RESULT 8

LOCUS AX161822

DEFINITION Sequence 5150 from Patent WO0140521.

ACCESSION AX161822

VERSION AX161822.1 GI:14543153

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 51)

AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eukaryota; Eutheria; Primates; Catarrhini; Homnidae; Homo. Shimkets, R.A. and Leach, M.

TITLE Nucleic acids containing single nucleotide polymorphisms and methods of use thereof

JOURNAL Patent: WO 0140521-A 5150 07-JUN-2001; Curagen Corporation (US)

FEATURES

source

1..51

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/note="2 of 2 allelic variants (5149 is other entry)"

misc\_feature

26

Accession number c943984006"

BASE COUNT 14 a 13 c 13 g 11 t

ORIGIN

Alignment Scores:

Pred. No.: 5.7e+03 Length: 51

Score: 47.00 Matches: 8

Percent Similarity: 72.73% Conservative: 0

Best Local Similarity: 72.73% Mismatches: 3

Query Match: 10.63% Indels: 0

DB: 6 Gaps: 0

US-09-599-087b-5 (1-81) x AX161822 (1-51)

OY 39 LeucysCysHisArgValProSerProAsnSer 49

DB 19 TTGTGCTGTGAACGACGACCGTCAGAGATTCC 51

RESULT 9

LOCUS AR077574

DEFINITION Sequence 289 from patent US 5962255.

ACCESSION AR077574

VERSION AR077574.1 GI:10004320

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 48)

AUTHORS Griffiths, A.David., Williams, S.Cameron., Waterhouse, P.Michael., Smith, A., Winter, G.Paul., Johnson, K.Stuart. and Nissim, A., John.Hammond.

TITLE Methods for producing recombinant vectors

JOURNAL Patent: US 5962255-A 289 05-OCT-1999;

FEATURES

source

1..48

/organism="unknown"

BASE COUNT 8 a 18 c 11 g 11 t

ORIGIN

Alignment Scores:

Pred. No.: 5.95e+03 Length: 48

Score: 46.50 Matches: 9

Percent Similarity: 66.67% Conservative: 3

Best Local Similarity: 50.00% Mismatches: 3

Query Match: 10.52% Indels: 3

Query Match:	10.41%	Indels:	0
DB:	6	Gaps:	0

US-09-599-087B-5 (1-81) x AX025094 (1-71)

25 LysArgArgProAlaLysAlaIlePseRSLyArgArgThrArgLeuCys 42  
:::|||||:::|||||  
Db 19 CGTGGCGCCCTCTGTCTCGGGTTCTGGAGACGCGTGTGAACATATGC 66

RESULT 12

LOCUS	171481	72 bp	DNA	linear	PAT 03-APR-1998
DEFINITION	Sequence 22 from patent US 5681942.				
ACCESSION	171481				
VERSION	171481.1	GI:3007616			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				

REFERENCE

1 (bases 1 to 72)  
Buchwald,M., Stratdee,C.A., Wevrick,R. and Mathew,C.George.Porter.  
TITLE  
Fancconi Anemia Type C gene  
JOURNAL  
Patent: US 5681942-A 22 28-OCT-1997;  
FEATURES

source	1..72	location/Qualifiers
BASE COUNT	12 a 16 c 13 g 31 t	
ORIGIN	/organism="unknown"	

Alignment Scores:

Pred. NO.:	1.03e+04	Length:	72
Score:	46.00	Matches:	7
Percent Similarity:	76.92%	Conservative:	3
Best local Similarity:	53.85%	Mismatches:	3
Query Match:	10.41%	Indels:	0
DB:	6	Gaps:	0

US-09-599-087B-5 (1-81) x 171481 (1-72)

41 CysH1ArgValProSerProAsnSerThrsAsnLeuLys 53  
|||||:::|||||  
Db 71 TGCCACAGATGGAATCCAAAGAGCATGAACATTAG 33

RESULT 13

HSIGRAB01	78 bp	mRNA	linear	PRI 04-AUG-1992
LOCUS				
DEFINITION	Human BL Ly66 mRNA for abnormal Ig kappa light chain (1).			
ACCESSION	X54166			
VERSION	X54166.1	GI:33218		
KEYWORDS	Ig kappa light chain; Ig light chain; immunoglobulin.			
SOURCE	Homo sapiens.			
ORGANISM	Homo sapiens			

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 78)  
COGNE,M.C.C.  
Direct Submission  
Submitted (31-JUL-1990) 'Cogne M.C.C.', CNRS URA 1172, Lab.  
Immunologie Molculaire, Faculte des Sciences, F-86022 Poitiers,  
france

2 (bases 1 to 78)  
Cogne,M., Mounlt,S., Mahdl,T., Preud'homme,J.L., Nau,F. and  
Guglielmi,P.  
Production of an abnormal mu chain with a shortened VHIV subgroup  
variable region in a Burkitt's lymphoma cell line  
Mol. Immunol. 27 (9), 929-934 (1990)  
31015095  
2120579

location/Qualifiers

1..78  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="Ly66"

	mRNA	/cell_type="B-cell" /tissue_type="Burkitt lymphoma"
exon		1..78 /note="abnormal Ig L chain 0.85kb transcript" /evidence=experimental 14..68 /note="v kappa 1 leader peptide exon"
exon		/number=1 69..>78 /note="C kappa exon, corresponds to Km3 allele sequence"
BASE COUNT	13 a	24 c 23 g 18 t
ORIGIN		/number=2
Alignment Scores:		
Pred. No.:	1.12e+04	Length: 78
Score:	46.00	Matches: 10
Percent Similarity:	76.47%	Conservative: 3
Best Local Similarity:	58.82%	Mismatches: 4
Query Match:	10.41%	Indels: 0
DB:	9	Gaps: 0
US-09-599-087B-5 (1-81) x HSIGKAB01 (1-78)		
OY	1 MetArgLeuValIeuSerIeuLeuCySIIleuLeuCyPhe 17	
Dn	14 ATGGACATGAGAGTCTCGTCAGCTCGTGggGgCTCTGTCTGTTTC 64	
RESULT 14		
LOCUS	AR005307	69 bp DNA linear PAT 04-DEC-1998
DEFINITION	Sequence 37 from patent US 5747660.	
ACCESSION	AR005307	
VERSION	AR005307.1	GI:3966186
KEYWORDS		
SOURCE	.	
ORGANISM	Unknown.	
REFERENCE	Unknown. Unclassified.	
AUTHORS	1 (bases 1 to 69) Orlicky,D.J.	
TITLE	Nucleic acid encoding prostaglandin F.sub.2.alpha. receptor regulatory protein	
JOURNAL	Patent: US 5747660-A 37 05-MAY-1998;	
FEATURES	Location/Qualifiers	
source	1..69 /organism="unknown"	
BASE COUNT	6 a 25 c 31 g 7 t	
ORIGIN		
Alignment Scores:		
Pred. No.:	1.1e+04	Length: 69
Score:	45.50	Matches: 10
Percent Similarity:	54.55%	Conservative: 2
Best Local Similarity:	45.45%	Mismatches: 7
Query Match:	10.29%	Indels: 3
DB:	6	Gaps: 1
US-09-599-087B-5 (1-81) x AR005307 (1-69)		
OY	25 LysAlaArgProAla-----LysAlaTrpSerGlyArgArgThrArgLeuCysCys 41	
Dn	3 AGGCCGAGAGTCCTCCGCCGCCGAGCATGGGGCCCCGCCGAGCCGCTGCTGC 62	
OY	42 HisArg 43	
Db	63 TGGCGC 68	
RESULT 15		
LOCUS	AR077518	39 bp DNA linear PAT 31-AUG-2000
DEFINITION	Sequence 233 from patent US 5962255.	
ACCESSION	AR077518	
VERSION	AR077518.1	GI:10004264

```

KEYWORDS
SOURCE
ORGANISM
AUTHORS
REFERENCE
TITL
JOURNAL
FEATURES
source
BASE COUNT
ORIGIN
Alignment Scores:
Pred. No.:
Score:
Percent Similarity:
Best Local Similarity:
Query Match:
DB:

Unknown.
Unknown.
Unclassified.
1 (bases 1 to 39)
Griffiths, A. David., Williams, S. Cameron., Waterhouse, P. Michael., and
Missim, A., Winter, G. Paul., Johnson, K. Stuart., and
Smith, A. John. Hammond.
Methods for producing recombinant vectors
Patent: US 5962255-A 233 05-Oct-1999;
Location/Qualifiers
1..39
"/organism="unknown"
7 a 13 c 10 g 9 t
Length: 39
Matches: 8
Conservative: 4
Mismatches: 0
Indels: 0
Gaps: 0

US-599-087B-5 (1-81) x AR077518 (1-39)
Oy 32 TrpserGIyArGArGTThrArGleuCYsCYsHIsArg 43
||||| |||||||
TGCTCCCTCCGCGCGAATACCAAGGCTGCTGTCAATAGG 38

```

```
Search completed: May 8, 2003, 23:26:38
Job time: 1229 secs
```

	PAT	31-AUG-2000
RESULT 15		
AR077518		
LOCUS	39 bp.	DNA
DEFINITION	Sequence 233 from patent US 596225.	linear
ACCESSION	AR077518	
VERSION	AR077518.1 GI:10004264	

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 21:46:33 ; Search time 21 Seconds  
(without alignments)  
354.956 Million cell updates/sec

Title: US-09-599-087b-5

Perfect score: 442

Sequence: 1 MRLVLSLLCILLCSIF.....PCKLEPRMWWGALPQV 81

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 349150 seqs, 92025710 residues

Total number of hits satisfying chosen parameters: 44737

Minimum DB seq length: 25  
Maximum DB seq length: 80

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PC1\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61.5	13.9	76	9	US-10-178-213-128
2	61.5	13.9	76	9	US-10-178-213-131
3	58.5	13.2	77	9	US-10-097-065-142
4	58	13.1	53	9	US-09-981-876-255
5	58	13.1	53	9	US-09-148-545-255
6	55.5	12.6	52	9	US-09-974-879-233
7	55	12.4	80	9	US-10-097-065-432
8	53.5	12.1	44	10	US-09-864-761-46147
9	53	12.0	61	10	US-09-815-242-4938
10	52.5	11.9	59	9	US-09-809-391-680
11	52.5	11.9	62	10	US-09-739-254-108
12	52.5	11.9	62	10	US-09-904-615-108
13	51.5	11.7	35	9	US-10-097-065-610
14	51	11.5	35	9	US-09-291-925A-1
15	51	11.5	47	9	US-09-809-391-523
16	49.5	11.2	80	9	US-10-178-213-89
17	49.5	11.2	73	9	US-09-864-761-40923
18	49.5	11.2	73	9	US-10-079-854-174
19	49.5	11.2	73	10	US-09-764-878-174

20	48	10.9	61	9	US-10-259-520-9	Sequence 9, App1
21	48	10.9	61	10	US-09-263-811-9	Sequence 9, App1
22	48	10.9	62	9	US-10-091-504-839	Sequence 839, App
23	48	10.9	62	10	US-09-764-869-839	Sequence 839, App
24	48	10.9	66	9	US-10-163-547-20	Sequence 20, App1
25	47.5	10.7	44	10	US-09-864-761-40752	Sequence 40752, A
26	47.5	10.7	69	9	US-09-910-009A-65	Sequence 65, App1
27	47.5	10.7	69	9	US-09-910-009A-219	Sequence 219, App
28	47.5	10.7	79	9	US-10-091-504-889	Sequence 889, App
29	47.5	10.7	79	10	US-09-764-869-889	Sequence 889, App
30	47	10.6	32	9	US-09-291-925A-6	Sequence 6, App1
31	47	10.6	72	9	US-09-898-659-29	Sequence 29, App1
32	46.5	10.5	32	9	US-09-983-802-207	Sequence 207, App
33	46.5	10.5	57	10	US-09-802-094-6	Sequence 6, App1
34	46.5	10.5	64	9	US-09-736-457-1830	Sequence 1830, Ap
35	46.5	10.5	64	9	US-09-902-941-1830	Sequence 1830, Ap
36	46.5	10.5	64	9	US-09-849-626-1830	Sequence 1830, Ap
37	46.5	10.5	64	9	US-10-017-754-1830	Sequence 1830, Ap
38	46.5	10.5	68	9	US-10-072-349-153	Sequence 153, App
39	46.5	10.5	68	10	US-09-764-865-153	Sequence 153, App
40	46.5	10.5	71	10	US-09-925-300-1861	Sequence 1861, Ap
41	46.5	10.5	73	8	US-08-424-550B-32	Sequence 32, App1
42	46.5	10.5	76	10	US-09-925-301-1187	Sequence 1187, Ap
43	46	10.4	40	10	US-09-864-761-45258	Sequence 45258, A
44	46	10.4	61	9	US-10-012-542-486	Sequence 486, App
45	46	10.4	61	10	US-09-867-550-1642	Sequence 1642, App

#### ALIGNMENTS

RESULT 1  
US-10-178-213-128  
; Sequence 128, Application US/10178213  
; Publication No. US20030041348A1  
GENERAL INFORMATION:  
; APPLICANT: Simmons, Carl R.  
; APPLICANT: Navarro Acevedo, Pedro A.  
; APPLICANT: Harwell, Leslie  
; APPLICANT: Cahoon, Rebecca  
; APPLICANT: McCutchen, Billy Fred  
; APPLICANT: Lu, Albert  
; APPLICANT: Herrmann, Rafael  
; APPLICANT: Wong, James  
; TITLE OF INVENTION: Defensin Polynucleotides and Methods of  
; TITLE OF INVENTION: Use  
; FILE REFERENCE: 35718/246703  
; CURRENT APPLICATION NUMBER: US/10/178,213  
; CURRENT FILING DATE: 2002-06-21  
; PRIOR APPLICATION NUMBER: 60/300,152  
; PRIOR FILING DATE: 2001-06-22  
; PRIOR APPLICATION NUMBER: 60/300,241  
; PRIOR FILING DATE: 2001-06-22  
; NUMBER OF SEQ ID NOS: 469  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 128  
; LENGTH: 76  
; TYPE: PRT  
; ORGANISM: Nicotiana benthamiana  
US-10-178-213-128  
Query Match 13.9%; Score 61.5; DB 9; Length 76;  
Best Local Similarity 33.3%; Pred. No. 14;  
Matches 14; Conservative 9; Mismatches 10; Indels 9; Gaps 3;  
OY 10 LCILLCF-----SIFSTEGKR--RPAKAGGR-RTRLCH 42  
DB 11 ICVVLFCFLMFASEVQITEAKHCKGKPSKSMGKCPKRCNH 52  
RESULT 2  
US-10-178-213-131  
; Sequence 131, Application US/10178213

```

? PRIOR FILING DATE: 1997-12-19
? NUMBER OF SEQ ID NOS: 672
? SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 142
LENGTH: 77
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (77)
OTHER INFORMATION: Xaa equals stop translation
US-10-097-065-142

Query Match      13.2%; Score 58.5; DB 9; Length 77;
Best local similarity 32.8%; Pred. No. 30;
Matches 19; Conservative 5; Mismatches 17; Indels 17; Gaps 2

QY          4 LVLSLCLTLLCF-----SITSEGRK--RPAKWSGRRTRLCRRV 44
            |||::|||::|::|::|::|::|::|::|::|::|::|::|::|::|
Db           11 LVLGSVFMIILTIYWDNSAAPTSTCTPRSLGTGRRCRPDPTGGAGHGLRCRRV 68

RESULT 4
US-09-981-876-255
; Sequence 255, Application US/09981876
; Patent No. US2002016469A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 70 Human Secreted Proteins
; FILE REFERENCE: P2001P1
; CURRENT APPLICATION NUMBER: US/09/981,876
; CURRENT FILING DATE: 2001-10-19
; PRIOR APPLICATION NUMBER: 09/148,545
; PRIOR FILING DATE: 1998-09-04
; PRIOR APPLICATION NUMBER: 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,333
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/038,621
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,161
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,626
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,334
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,336
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/040,163
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: 60/047,615
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,600
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,597
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,502
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,633
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,583
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,617
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,618
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,503
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,592
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,581
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: 60/047,584

```

[illegible]



1	EARLIER APPLICATION NUMBER: 1997-05-23	5888
2	EARLIER FILING DATE: 1997-05-23	5888
3	EARLIER APPLICATION NUMBER: 60/04/7, 6113	5888
4	EARLIER FILING DATE: 1997-05-23	5888
5	EARLIER APPLICATION NUMBER: 60/04/7, 5888	5888
6	EARLIER FILING DATE: 1997-05-23	5888
7	EARLIER APPLICATION NUMBER: 60/04/7, 5988	5888
8	EARLIER FILING DATE: 1997-05-23	5888
9	EARLIER APPLICATION NUMBER: 60/04/7, 6112	5888
10	EARLIER FILING DATE: 1997-05-23	5888
11	EARLIER APPLICATION NUMBER: 60/04/7, 6338	5888
12	EARLIER FILING DATE: 1997-05-23	5888
13	EARLIER APPLICATION NUMBER: 60/04/7, 6018	5888
14	EARLIER FILING DATE: 1997-04-11	5888
15	EARLIER APPLICATION NUMBER: 60/04/3, 5658	5888
16	EARLIER FILING DATE: 1997-04-11	5888
17	EARLIER APPLICATION NUMBER: 60/04/3, 3148	5888
18	EARLIER FILING DATE: 1997-04-11	5888
19	EARLIER APPLICATION NUMBER: 60/04/3, 6748	5888
20	EARLIER FILING DATE: 1997-04-11	5888
21	EARLIER APPLICATION NUMBER: 60/04/3, 6658	5888
22	EARLIER FILING DATE: 1997-04-11	5888
23	EARLIER APPLICATION NUMBER: 60/04/3, 3118	5888
24	EARLIER FILING DATE: 1997-04-11	5888
25	EARLIER APPLICATION NUMBER: 60/04/3, 6718	5888
26	EARLIER FILING DATE: 1997-04-11	5888
27	EARLIER APPLICATION NUMBER: 60/04/3, 3138	5888
28	EARLIER FILING DATE: 1997-04-11	5888
29	EARLIER APPLICATION NUMBER: 60/04/3, 6728	5888
30	EARLIER FILING DATE: 1997-04-11	5888
31	EARLIER APPLICATION NUMBER: 60/04/3, 3158	5888
32	EARLIER FILING DATE: 1997-04-11	5888
33	EARLIER APPLICATION NUMBER: 60/048, 9748	5888
34	EARLIER FILING DATE: 1997-06-06	5888
35	EARLIER APPLICATION NUMBER: 60/056, 8868	5888
36	EARLIER FILING DATE: 1997-08-22	5888
37	EARLIER APPLICATION NUMBER: 60/056, 8778	5888
38	EARLIER FILING DATE: 1997-08-22	5888
39	EARLIER APPLICATION NUMBER: 60/056, 8888	5888
40	EARLIER FILING DATE: 1997-08-22	5888
41	EARLIER APPLICATION NUMBER: 60/056, 6628	5888
42	EARLIER FILING DATE: 1997-08-22	5888
43	EARLIER APPLICATION NUMBER: 60/056, 8728	5888
44	EARLIER FILING DATE: 1997-08-22	5888
45	EARLIER APPLICATION NUMBER: 60/056, 9038	5888
46	EARLIER FILING DATE: 1997-08-22	5888
47	EARLIER APPLICATION NUMBER: 60/056, 8888	5888
48	EARLIER FILING DATE: 1997-08-22	5888
49	EARLIER APPLICATION NUMBER: 60/056, 8948	5888
50	EARLIER FILING DATE: 1997-08-22	5888
51	EARLIER APPLICATION NUMBER: 60/056, 9118	5888
52	EARLIER FILING DATE: 1997-08-22	5888
53	EARLIER APPLICATION NUMBER: 60/056, 6368	5888
54	EARLIER FILING DATE: 1997-05-23	5888

```

1 EARLIER FILING DATE: 1997-08-22
2 EARLIER APPLICATION NUMBER: 60/056,874
3
4 EARLIER FILING DATE: 1997-08-22
5 EARLIER APPLICATION NUMBER: 60/056,910
6
7 EARLIER FILING DATE: 1997-08-22
8 EARLIER APPLICATION NUMBER: 60/056,864
9
10 EARLIER FILING DATE: 1997-08-22
11 EARLIER APPLICATION NUMBER: 60/056,631
12
13 EARLIER FILING DATE: 1997-08-22
14 EARLIER APPLICATION NUMBER: 60/056,845
15
16 EARLIER FILING DATE: 1997-08-22
17 EARLIER APPLICATION NUMBER: 60/056,892
18
19 EARLIER FILING DATE: 1997-08-22
20 EARLIER APPLICATION NUMBER: 60/047,595
21
22 EARLIER FILING DATE: 1997-05-23
23 EARLIER APPLICATION NUMBER: 60/057,761
24
25 EARLIER FILING DATE: 05-Sep-1997
26 EARLIER APPLICATION NUMBER: 60/047,559
27
28 EARLIER FILING DATE: 1997-05-23
29 EARLIER APPLICATION NUMBER: 60/047,588
30
31 EARLIER FILING DATE: 1997-05-23
32 EARLIER APPLICATION NUMBER: 60/047,585
33
34 EARLIER FILING DATE: 1997-05-23
35 EARLIER APPLICATION NUMBER: 60/047,586
36
37 EARLIER FILING DATE: 1997-05-23
38 EARLIER APPLICATION NUMBER: 60/047,590
39
40 EARLIER FILING DATE: 1997-05-23
41 EARLIER APPLICATION NUMBER: 60/047,594
42
43 EARLIER FILING DATE: 1997-05-23
44 EARLIER APPLICATION NUMBER: 60/047,589
45
46 EARLIER FILING DATE: 1997-05-23
47 EARLIER APPLICATION NUMBER: 60/047,593
48
49 EARLIER FILING DATE: 1997-05-23
50 EARLIER APPLICATION NUMBER: 60/047,614
51
52 EARLIER FILING DATE: 1997-05-23
53 EARLIER APPLICATION NUMBER: 60/043,578
54
55 EARLIER FILING DATE: 1997-04-11
56 EARLIER APPLICATION NUMBER: 60/043,576
57
58 EARLIER FILING DATE: 1997-04-11
59 EARLIER APPLICATION NUMBER: 60/047,501
60
61 EARLIER FILING DATE: 1997-05-23
62 EARLIER APPLICATION NUMBER: 60/043,670
63
64 EARLIER FILING DATE: 1997-04-11
65 EARLIER APPLICATION NUMBER: 60/056,632
66
67 EARLIER FILING DATE: 1997-08-22
68 EARLIER APPLICATION NUMBER: 60/056,664
69
70 EARLIER FILING DATE: 1997-08-22
71 EARLIER APPLICATION NUMBER: 60/056,876
72
73 EARLIER FILING DATE: 1997-08-22
74 EARLIER APPLICATION NUMBER: 60/056,881
75
76 EARLIER FILING DATE: 1997-08-22
77 EARLIER APPLICATION NUMBER: 60/056,909
78
79 EARLIER FILING DATE: 1997-08-22
80 EARLIER APPLICATION NUMBER: 60/056,875
81
82 EARLIER FILING DATE: 1997-08-22
83 EARLIER APPLICATION NUMBER: 60/056,862
84
85 EARLIER FILING DATE: 1997-08-22
86 EARLIER APPLICATION NUMBER: 60/056,887
87
88 EARLIER FILING DATE: 1997-08-22
89 EARLIER APPLICATION NUMBER: 60/056,908
90
91 EARLIER FILING DATE: 1997-08-22
92 EARLIER APPLICATION NUMBER: 60/048,964
93
94 EARLIER FILING DATE: 1997-06-06
95 EARLIER APPLICATION NUMBER: 60/057,650
96
97 EARLIER FILING DATE: 1997-09-05
98 EARLIER APPLICATION NUMBER: 60/056,884
99
100 EARLIER FILING DATE: 1997-08-22
101 NUMBER OF SEQ ID NOS: 280
102
103 SOFTWARE: PatentIn Ver. 2.0
104
105 SEQ ID NO 255
106
107 LENGTH: 53

```

Query Match	13.1%; Score 58; DB 9; Length 53;
-------------	-----------------------------------

```

Best Local Similarity 50.0%; Pred. No. 23;
Matches 12; Conservative 6; Mismatches 6; Indels 0; Gaps 0.

      3  LTVSSLCITLLDLCFSIFSTEGKR 26
      |||:|:| : | : | : | : |
Db . 19 LLLLSLAPMLVLCYMRMPPTKGR 42

```

RESULT 6  
US-09-974-879-233  
; Sequence 233, Application US/09974879  
; Publication No. US20030028003A1  
; GENERAL INFORMATION:

```

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 125 Human Secreted Proteins

```

FILE REFERENCE: PZ020P2  
CURRENT APPLICATION NUMBER: US/09/974.879

CURRENT FILING DATE: 2001-10-12  
CURRENT FILING NUMBER: 001-10-12

;; PRIOR APPLICATION NUMBER: US 00/239,893  
; PRIOR FILING DATE: 2000-10-13

;  
; PRIOR APPLICATION NUMBER: US 09/818,683  
;  
; PRIOR FILING DATE: 2001-03-28  
;

PRIOR APPLICATION NUMBER: US 09/305,736  
PRIOR FILING DATE: 1999-05-05

PRIOR APPLICATION NUMBER: PCT/US98/23435

;; PRIOR FILING DATE: 1998-11-04  
;; PRIOR APPLICATION NUMBER: US 60/064,911

; PRIOR FILING DATE: 1997-11-07  
 ; PRIOR APPLICATION NUMBER: US 60/064,912

PRIOR FILING DATE: 1997-11-07  
PRIOR APPLICATION NUMBER: US 60/064,983

PRIOR FILING DATE: 1997-11-07

PRIOR FILING DATE: 1997-11-07

;; PRIOR APPLICATION NUMBER: US 60/064,988  
; PRIOR FILING DATE: 1997-11-07

PRIOR APPLICATION NUMBER: US 60/064, 987  
PRIOR FILING DATE: 1997-11-07

PRIOR APPLICATION NUMBER: US 60/064,908  
PRIOR FILING DATE: 1997-11-07

PRIOR APPLICATION NUMBER: US 60/064,984

;; PRIOR FILING DATE: 1997-11-07  
;; PRIOR APPLICATION NUMBER: US 60/064,985

; PRIOR FILING DATE: 1997-11-07  
 ; PRIOR APPLICATION NUMBER: US 60/066,094

PRIOR FILING DATE: 1997-11-17

; PRIOR APPLICATION NUMBER: US 60/088,400  
 ; PRIOR FILING DATE: 1997-11-17

;; PRIOR APPLICATION NUMBER: US 60/066,089  
;; PRIOR FILING DATE: 1997-11-17

; PRIOR APPLICATION NUMBER: US 60/066,095  
 ; PRIOR FILING DATE: 1997-11-17

PRIOR APPLICATION NUMBER: US 60/066,090

; PRIOR FILING DATE: 1997-11-1/  
; NUMBER OF SEQ ID NOS: 611

```

; SOFTWARE: PatentIn Ver. 2.0
; SEO ID NO 233

```

```

; LENGTH: 52
; TYPE: PRT

```

ORGANISM: Homo sapiens

US-09-9/4-8/9-233

Query Match	12.68;	Score	55.5
Best Local Similarity	34.08;	Pred. No.	

Matches 17; Conservative 4; Mismatch

QY 8 SLLCILLFSTFSTEGKRRPAKAWSGRRTRLCC

```
Db      6 STCLVLFCSVCSEASCOSTVQMGAAARV--
```

RESULT 7  
ITS-10-097-065-433

C  
U  
H  
C  
  
C  
/  
  
C  
C  
C  
  
N  
C  
N

Query Match	12.6%	Score 55.5	DB 9	Length 52
Best Local Similarity	34.0%	Pred. No. 40		
Matches 17, Conservative	4	Mismatches	24	Indels 5
				Gaps 2

QY 8 SLICILLLCFSTIEGRRPAAWMSGRRTRLCCHRVSPNSINL--KGH 55  
+ + + + +  
Db 6 STICVLISPCSVCSVEASQSTVQMGGAAAR--GVPEDMRSRDEOGKH 52  
+ + + + +

RESULT 7  
US-10-097-065-432

```
; Sequence 432, Application US/10097065
; Publication No. US20030055236A1
; GENERAL INFORMATION:
; APPLICANT: Moore, Paul A. et al.
; TITLE OF INVENTION: 110 Human Secreted Proteins
; FILE REFERENCE: P2021P1
; CURRENT APPLICATION NUMBER: US/10/097,065
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: PCT/US98/27059
; PRIOR FILING DATE: 1998-12-17
; PRIOR APPLICATION NUMBER: 60/070,923
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,007
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,057
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,006
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,369
; PRIOR FILING DATE: 1997-12-19
; PRIOR APPLICATION NUMBER: 60/068,367
; PRIOR FILING DATE: 1997-12-19
; PRIOR APPLICATION NUMBER: 60/068,368
; PRIOR FILING DATE: 1997-12-19
; PRIOR APPLICATION NUMBER: 60/068,169
; PRIOR FILING DATE: 1997-12-19
; PRIOR APPLICATION NUMBER: 60/068,053
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,064
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,054
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,008
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/068,365
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 432
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-097-065-432
;
; Query Match 12.4%; Score 55; DB 9; Length 80;
; Best Local Similarity 33.3%; Pred. No. 73;
; Matches 16; Conservative 7; Mismatches 23; Indels 2; Gaps 1;
```

```
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 46147
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC023890.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4
; OTHER INFORMATION: SWISSPROT HIT: P45816, EVALU 4.80e+00
; OTHER INFORMATION: EST_HUMAN HIT: BE545503.1, EVALU 7.00e-19
; US-09-864-761-46147
;
; Query Match 12.1%; Score 53.5; DB 10; Length 44;
; Best Local Similarity 37.1%; Pred. No. 55;
; Matches 13; Conservative 6; Mismatches 15; Indels 1; Gaps 1;
```

```

; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4938
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-4938

Query Match          12.0%; Score 53; DB 10; Length 61;
Best Local Similarity 40.0%; Pred. No. 88;
Matches 14; Conservative 4; Mismatches 15; Indels 2; Gaps 1;

QY 26 RPAKAMSGRRTRLCCHRVSPSTNLKGHHVRLC 60
DB 11 KRPAKHSSTAYTR-CERCGRPHSVYRKFLCRIC 43

RESULT 10
US-09-809-391-680
; Sequence 680, Application US/09809391
; Publication No. US20030049618A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P2
; CURRENT APPLICATION NUMBER: US/09/809,391
; CURRENT FILING DATE: 2001-03-16
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 761
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 680
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-809-391-680

Query Match          11.9%; Score 52.5; DB 9; Length 59;
Best Local Similarity 33.3%; Pred. No. 96;
Matches 14; Conservative 7; Mismatches 18; Indels 3; Gaps 2;

QY 27 RPAKAMSGRRTRLCCHRVSPSTNLKGHHVRLCCKPCKLEP 67
DB 13 RPSRGWSSRTSVSRHRDTEHTSSRSKGTSLQLI--CKSEP 52

RESULT 11
US-09-739-254-108
; Sequence 108, Application US/09739254
; Patent No. US20010021700A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 49 Human Secreted Proteins
; FILE REFERENCE: P2032P1
; CURRENT APPLICATION NUMBER: US/09/739,254
; CURRENT FILING DATE: 2000-12-19
; EARLIER APPLICATION NUMBER: 09/511,554
; EARLIER FILING DATE: 2000-02-23
```

```

; EARLIER APPLICATION NUMBER: PCT/US99/19330
; EARLIER FILING DATE: 1999-08-24
; EARLIER APPLICATION NUMBER: 60/097,917
; EARLIER FILING DATE: 1998-08-25
; EARLIER APPLICATION NUMBER: 60/098,634
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 108
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (62)
; OTHER INFORMATION: Xaa equals stop translation
US-09-739-254-108

Query Match          11.9%; Score 52.5; DB 10; Length 62;
Best Local Similarity 29.3%; Pred. No. 1e+02;
Matches 17; Conservative 11; Mismatches 23; Indels 7; Gaps 3;

QY 10 ICILLCFISF--TEGKR---PAKAMSGRRTRLCCHRVSP--NSTNLKGHHVRLC 60
DB 1 MCLLLMLTTFORTSGALRRGGLSSPAKAMRSPSVYSTQTPPMSTGTLRGLSGANC 58

RESULT 12
US-09-904-615-108
; Sequence 108, Application US/09904615
; Patent No. US2002026040A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 49 Human Secreted Proteins
; FILE REFERENCE: P2032P1
; CURRENT APPLICATION NUMBER: US/09/904,615
; CURRENT FILING DATE: 2001-07-16
; PRIOR APPLICATION NUMBER: 09/511,554
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: 60/097,917
; PRIOR FILING DATE: 1998-08-25
; PRIOR APPLICATION NUMBER: 60/098,634
; PRIOR FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 108
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (62)
; OTHER INFORMATION: Xaa equals stop translation
US-09-904-615-108

Query Match          11.9%; Score 52.5; DB 10; Length 62;
Best Local Similarity 29.3%; Pred. No. 1e+02;
Matches 17; Conservative 11; Mismatches 23; Indels 7; Gaps 3;

QY 10 ICILLCFISF--TEGKR---PAKAMSGRRTRLCCHRVSP--NSTNLKGHHVRLC 60
DB 1 MCLLLMLTTFORTSGALRRGGLSSPAKAMRSPSVYSTQTPPMSTGTLRGLSGANC 58

RESULT 13
US-10-097-065-610
; Sequence 610, Application US/10097065
; Publication No. US20030055236A1
; GENERAL INFORMATION:
; APPLICANT: Moore, Paul A. et al.
; TITLE OF INVENTION: 110 Human Secreted Proteins
; FILE REFERENCE: P2021P1
; CURRENT APPLICATION NUMBER: US/10/097,065
```

RESULT 14  
 US-09-291-925A-1  
 : Sequence 1, Application US/09291925A  
 : Patent No. US20020155525A1  
 : GENERAL INFORMATION:  
 : APPLICANT: Ashkenazi, A.  
 : APPLICANT: Berman, P.  
 : APPLICANT: Brousseau, D.  
 : APPLICANT: Etcheverry, T.  
 : TITLE OF INVENTION: SECRETION OF GLYCOSYLATION MUTANTS  
 : FILE REFERENCE: P1055R1  
 : CURRENT APPLICATION NUMBER: US/09/291,925A  
 : CURRENT FILING DATE: 1999-04-14  
 : PRIOR APPLICATION NUMBER: US 60/082,002  
 : PRIOR FILING DATE: 1998-04-16  
 : PRIOR APPLICATION NUMBER: US 60/123,522  
 : PRIOR FILING DATE: 1999-03-08  
 : NUMBER OF SEQ ID NOS: 13  
 : SEQ ID NO 1  
 : LENGTH: 35  
 : TYPE: PRT  
 : ORGANISM: Homo sapiens  
 : US-09-291-925A-1.

```
Search completed: May  8, 2003, 21:49:15
Job time : 23 secs
```

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 21:45:48 ; Search time 18 Seconds

(without alignments)  
132,403 Million cell updates/sec

Title: US-09-599-087b-5

Perfect score: 442  
Sequence: 1 MRLIVSSLCITLILCFISIF.....PCKLEPPRLWVPGALPOV 81

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 2942292 residues

Total number of hits satisfying chosen parameters: 46248

Minimum DB seq length: 25  
Maximum DB seq length: 80

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents, AA: \*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep: \*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep: \*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep: \*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep: \*  
5: /cgn2\_6/ptodata/2/1aa/PCITUS.COMB.pep: \*  
6: /cgn2\_6/ptodata/2/1aa/Backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52.5	11.9	59	4	US-09-149-476-680
2	52.5	11.9	74	1	US-08-543-238-5
3	52.5	11.9	74	1	US-08-420-528-5
4	51	11.5	42	4	US-08-974-549A-174
5	51	11.5	47	4	US-09-149-476-523
6	50	11.3	56	2	US-08-637-759B-123
7	50	11.3	56	3	US-08-871-355A-123
8	50	11.3	56	4	US-09-201-945-123
9	50	11.3	77	2	US-08-726-306A-174
10	49.5	11.2	77	1	US-08-443-568B-10
11	49.5	11.2	77	5	PCIT-US94-06997-10
12	48	10.9	36	6	5256770-8
13	48	10.9	40	6	5244676-2
14	48	10.9	61	3	US-08-743-975-9
15	47.5	10.7	68	1	US-07-689-693B-11
16	47	10.6	35	1	US-07-609-510B-1
17	46.5	10.5	35	1	US-08-427-640-10
18	46.5	10.5	32	4	US-09-227-357-207
19	46.5	10.5	45	4	US-08-891-271-8
20	46.5	10.5	69	4	US-09-228-302-7
21	46.5	10.5	73	4	US-08-469-260A-32
22	46	10.4	32	3	US-08-792-832A-48
23	46	10.4	66	4	US-09-383-586-15
24	45.5	10.3	80	4	US-09-188-930-150
25	45.5	10.3	80	4	US-09-188-930-300
26	45	10.2	28	3	US-08-848-580-10
27	45	10.2	28	5	PCIT-US92-04537-6

28	45	10.2	35	1	US-08-282-030-11
29	45	10.2	35	5	PCIT-US95-10219-11
30	45	10.2	75	2	US-08-783-385-2
31	45	10.2	75	2	US-08-924-838-8
32	44.5	10.1	38	3	US-08-504-538A-17
33	44.5	10.1	38	4	US-08-630-052-17
34	44.5	10.1	38	5	PCIT-US95-09307-17
35	44.5	10.1	46	4	US-08-632-511A-5
36	44.5	10.1	46	4	US-09-091-550A-11
37	44.5	10.1	46	4	US-09-488-200-5
38	44.5	10.1	70	4	US-09-006-428A-15
39	44	10.0	52	3	US-08-851-843A-184
40	44	10.0	52	4	US-08-974-549A-303
41	44	10.0	52	4	US-08-854-050-184
42	44	10.0	52	4	US-09-430-323-184
43	44	10.0	67	1	US-08-471-780C-126
44	44	10.0	67	1	US-08-467-282B-126
45	44	10.0	67	2	US-08-471-282A-126

# ALIGNMENTS

RESULT 1  
US-09-149-476-680  
Sequence 680, Application US/09149476  
Patent No. 6420526  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: 186 Human Secreted proteins  
FILE REFERENCE: P2002PI  
CURRENT APPLICATION NUMBER: US/09/149,476  
CURRENT FILING DATE: 1998-09-08  
EARLIER APPLICATION NUMBER: PCIT/US98/04493  
EARLIER FILING DATE: 1998-03-06  
EARLIER APPLICATION NUMBER: 60/040,162  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,333  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/038,621  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,626  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,334  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,336  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,163  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/047,600  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,615  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,597  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,502  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,633  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,583  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,617  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,618  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,503  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,592  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,581  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,584  
EARLIER FILING DATE: 1997-05-23

Sequence 11, Appl  
Sequence 11, Appl  
Sequence 2, Appl  
Sequence 8, Appl  
Sequence 17, Appl  
Sequence 17, Appl  
Sequence 17, Appl  
Sequence 5, Appl  
Sequence 11, Appl  
Sequence 5, Appl  
Sequence 15, Appl  
Sequence 184, App  
Sequence 184, App  
Sequence 303, App  
Sequence 184, App  
Sequence 184, App  
Sequence 126, App  
Sequence 126, App  
Sequence 126, App

[illegible]

EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056, 8894
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056, 9111
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056, 6366
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056, 8744
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056, 9110
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056, 8664
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056, 6311
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056, 8453
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/047, 5555
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047, 5599
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047, 5688
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047, 5855
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047, 5868
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047, 5900
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047, 5944
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047, 5899
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047, 5939
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047, 5101
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/043, 5787
EARLIER	FILING DATE:	1997-04-11
EARLIER	APPLICATION NUMBER:	60/043, 5767
EARLIER	FILING DATE:	1997-04-11
EARLIER	APPLICATION NUMBER:	60/047, 5010
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/043, 6707
EARLIER	FILING DATE:	1997-04-11
EARLIER	APPLICATION NUMBER:	60/056, 6322
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056, 6654
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056, 8767
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056, 8811
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056, 9009
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056, 8757
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056, 8622
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056, 8887
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056, 9088
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/048, 9644
EARLIER	FILING DATE:	1997-06-06
EARLIER	APPLICATION NUMBER:	60/057, 6500
EARLIER	FILING DATE:	1997-09-05
EARLIER	APPLICATION NUMBER:	60/056, 8844
EARLIER	FILING DATE:	1997-08-22

EARLIER APPLICATION NUMBER: 60/057,669  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/049,610  
EARLIER FILING DATE: 1997-06-13  
EARLIER APPLICATION NUMBER: 60/061,060  
EARLIER FILING DATE: 1997-10-02

Query Match 11.9%; Score 52.5; DB 4; Length 59;  
Best Local Similarity 33.3%; Pred. No. 11;  
Matches 14; Conservative 7; Mismatches 18; Indels 3; Gaps 2;

QY 27 RPAKMSGRTRLCGR-VSPNSINKGHVRCCKLP 67  
DB 13 RPSGRWSRSTSVSRHRTENTSSRSKTSGLI--CKSEP 52

RESULT 2  
US-08-543-238-5  
Sequence 5, Application US/08543238

PATENT No. 5607919  
GENERAL INFORMATION:  
APPLICANT: Bojsen, Kirsten  
APPLICANT: Kragh, Karsten M.  
APPLICANT: Mikkelsen, John D.  
TITLE OF INVENTION: Anti-Microbial Proteins  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sandoz Agro, Inc.  
STREET: 975 California Avenue  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/543,238  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Marcus-Wyner, Lynn  
REGISTRATION NUMBER: 34,869  
REFERENCE/DOCKET NUMBER: 137-1078/MA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/354-3588  
TELEFAX: 415/857-1125  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 74 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-543-238-5

Query Match 11.9%; Score 52.5; DB 1; Length 74;  
Best Local Similarity 25.4%; Pred. No. 15;  
Matches 17; Conservative 10; Mismatches 25; Indels 15; Gaps 2;

QY 12 ILILCFSTFSTEGK-----RRPAKMSG-----RRTRLCGRVSPNSINKGHH 56  
DB 8 LLLLLFVASELMVAEVGATCRKPSMYFGACFSPTNCKAKCNREDWNGKCLVGFK 67  
QY 57 VRLCKPC 63  
DB 68 CECORPC 74

RESULT 3  
US-08-420-526-5

Sequence 5, Application US/08420526  
Patent No. 5608151  
GENERAL INFORMATION:  
APPLICANT: Bojsen, Kirsten  
APPLICANT: Kragh, Karsten M.  
APPLICANT: Mikkelsen, John D.  
TITLE OF INVENTION: Anti-Microbial Proteins  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sandoz Agro, Inc.  
STREET: 975 California Avenue  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/420,526  
FILING DATE:  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Marcus-Wyner, Lynn  
REGISTRATION NUMBER: 34,869  
REFERENCE/DOCKET NUMBER: 137-1078/MA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/354-3588  
TELEFAX: 415/857-1125  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 74 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-420-526-5

Query Match 11.9%; Score 52.5; DB 1; Length 74;  
Best Local Similarity 25.4%; Pred. No. 15;  
Matches 17; Conservative 10; Mismatches 25; Indels 15; Gaps 2;

QY 12 ILILCFSTFSTEGK-----RRPAKMSG-----RRTRLCGRVSPNSINKGHH 56  
DB 8 LLLLLFVASELMVAEVGATCRKPSMYFGACFSPTNCKAKCNREDWNGKCLVGFK 67  
QY 57 VRLCKPC 63  
DB 68 CECORPC 74

RESULT 4  
US-08-974-549A-174  
Sequence 174, Application US/08974549A  
Patent No. 6166178  
GENERAL INFORMATION:  
APPLICANT: Cecch, Thomas R.  
APPLICANT: Lingner, Joachim  
APPLICANT: Nakamura, Toru  
APPLICANT: Chapman, Karen B.  
APPLICANT: Morin, Gregg B.  
APPLICANT: Harley, Calvin B.  
APPLICANT: Andrews, William H.  
TITLE OF INVENTION: Human telomerase Catalytic Subunit  
NUMBER OF SEQUENCES: 727  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Crew LLP  
STREET: Two Embarcadero Center, Eighth Floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA

QY 12 ILILCFSTFSTEGK-----RRPAKMSG-----RRTRLCGRVSPNSINKGHH 56  
DB 8 LLLLLFVASELMVAEVGATCRKPSMYFGACFSPTNCKAKCNREDWNGKCLVGFK 67  
QY 57 VRLCKPC 63  
DB 68 CECORPC 74



ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/974,549A  
FILING DATE: 19-NOV-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/724,643  
FILING DATE: 01-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/844,419  
FILING DATE: 18-APR-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/846,017  
FILING DATE: 25-APR-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/851,843  
FILING DATE: 06-MAY-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/854,050  
FILING DATE: 09-MAY-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/911,312  
FILING DATE: 14-AUG-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/912,951  
FILING DATE: 14-AUG-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/915,503  
FILING DATE: 14-AUG-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US97/17618  
FILING DATE: 01-OCT-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US97/17885  
FILING DATE: 01-OCT-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Apple, Randolph Ted  
REGISTRATION NUMBER: 36,429  
REFERENCE/DOCKET NUMBER: 015389-002610US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 174:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 42 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-974-549A-174

Query Match 11.5%; Score 51; DB 4; Length 42;  
Best Local Similarity 36.8%; Pred. NO. 12;  
Matches 14; Conservative 4; Mismatches 18; Indels 2; Gaps 1;

4 LVLSSLLCILLICFISFEGRKRRRAKAMSGRTRLCC 41  
1 LTTDYLFCLLLIYIF--KRSRRNEKTKERDFKIC 36

US-09-149-476-523  
Sequence 523 Application US/09149476  
Patent No. 6420526  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: 186 Human Secreted proteins  
FILE REFERENCE: pz002p1

CURRENT APPLICATION NUMBER: US/09/149,476  
CURRENT FILING DATE: 1998-09-08  
EARLIER APPLICATION NUMBER: PCT/US98/04493  
EARLIER FILING DATE: 1998-03-06  
EARLIER APPLICATION NUMBER: 60/040,162  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,333  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/038,621  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,626  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,334  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,336  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/040,163  
EARLIER FILING DATE: 1997-03-07  
EARLIER APPLICATION NUMBER: 60/047,600  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,615  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,597  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,502  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,633  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,583  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,617  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,618  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,503  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,592  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,581  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,584  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,500  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,587  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,492  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,598  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,613  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,582  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,596  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,612  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,632  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,601  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,580  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,568  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,314  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,569  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,311  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,671

```

1 EARLIER APPLICATION NUMBER: 60/047,590
2 EARLIER FILING DATE: 1997-05-23
3 EARLIER APPLICATION NUMBER: 60/047,594
4 EARLIER FILING DATE: 1997-05-23
5 EARLIER APPLICATION NUMBER: 60/047,589
6 EARLIER FILING DATE: 1997-05-23
7 EARLIER APPLICATION NUMBER: 60/047,593
8 EARLIER FILING DATE: 1997-05-23
9 EARLIER APPLICATION NUMBER: 60/047,614
10 EARLIER FILING DATE: 1997-05-23
11 EARLIER APPLICATION NUMBER: 60/043,578
12 EARLIER FILING DATE: 1997-04-11
13 EARLIER APPLICATION NUMBER: 60/043,576
14 EARLIER FILING DATE: 1997-04-11
15 EARLIER APPLICATION NUMBER: 60/047,501
16 EARLIER FILING DATE: 1997-05-23
17 EARLIER APPLICATION NUMBER: 60/043,670
18 EARLIER FILING DATE: 1997-04-11
19 EARLIER APPLICATION NUMBER: 60/056,632
20 EARLIER FILING DATE: 1997-08-22
21 EARLIER APPLICATION NUMBER: 60/056,664
22 EARLIER FILING DATE: 1997-08-22
23 EARLIER APPLICATION NUMBER: 60/056,876
24 EARLIER FILING DATE: 1997-08-22
25 EARLIER APPLICATION NUMBER: 60/056,881
26 EARLIER FILING DATE: 1997-08-22
27 EARLIER APPLICATION NUMBER: 60/056,909
28 EARLIER FILING DATE: 1997-08-22
29 EARLIER APPLICATION NUMBER: 60/056,875
30 EARLIER FILING DATE: 1997-08-22
31 EARLIER APPLICATION NUMBER: 60/056,862
32 EARLIER FILING DATE: 1997-08-22
33 EARLIER APPLICATION NUMBER: 60/056,887
34 EARLIER FILING DATE: 1997-08-22
35 EARLIER APPLICATION NUMBER: 60/056,908
36 EARLIER FILING DATE: 1997-08-22
37 EARLIER APPLICATION NUMBER: 60/048,964
38 EARLIER FILING DATE: 1997-06-06
39 EARLIER APPLICATION NUMBER: 60/057,650
40 EARLIER FILING DATE: 1997-09-05
41 EARLIER APPLICATION NUMBER: 60/056,884
42 EARLIER FILING DATE: 1997-08-22
43 EARLIER APPLICATION NUMBER: 60/057,669
44 EARLIER FILING DATE: 1997-09-05
45 EARLIER APPLICATION NUMBER: 60/049,610
46 EARLIER FILING DATE: 1997-06-13
47 EARLIER APPLICATION NUMBER: 60/061,060
48 EARLIER FILING DATE: 1997-10-02
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
8
```

ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/637,759B  
FILING DATE: 03-MAY-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB95/02875  
FILING DATE: 11-DEC-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: RPMS 101  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 873-8794  
TELEFAX: (404) 873-8795  
INFORMATION FOR SEQ ID NO: 123:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 56 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-08-637-759B-123

Query Match 11.3%; Score 50; DB 2; Length 56;  
Best Local Similarity 34.4%; Pred. No. 22;  
Matches 11; Conservative 4; Mismatches 13; Indels 4; Gaps 1;

Oy 41 CHRVSPNSTLKGHVRLCKPKLEPERLM 72  
Db 14 CORVTERLQSNLSRYHNR----CSLRSDVRYW 41

RESULT 7  
US-08-871-355A-123  
Sequence 123, Application US/08871355A  
Patent No. 6015669  
GENERAL INFORMATION:  
APPLICANT: David William Holden  
TITLE OF INVENTION: Identification of Genes  
NUMBER OF SEQUENCES: 501  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center  
STREET: 1201 West Peachtree Street  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/871,355A  
FILING DATE: 09-JUN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB95/02875  
FILING DATE: 11-DEC-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: RPMS 101 CON

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 873-8794  
TELEFAX: (404) 873-8795  
INFORMATION FOR SEQ ID NO: 123:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 56 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-08-871-355A-123

Query Match 11.3%; Score 50; DB 3; Length 56;  
Best Local Similarity 34.4%; Pred. No. 22;  
Matches 11; Conservative 4; Mismatches 13; Indels 4; Gaps 1;

Oy 41 CHRVSPNSTLKGHVRLCKPKLEPERLM 72  
Db 14 CORVTERLQSNLSRYHNR----CSLRSDVRYW 41

RESULT 8  
US-09-201-945-123  
Sequence 123, Application US/09201945  
Patent No. 6342215  
GENERAL INFORMATION:  
APPLICANT: David William Holden  
TITLE OF INVENTION: Identification of Genes  
NUMBER OF SEQUENCES: 501  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center  
STREET: 1201 West Peachtree Street  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/201,945  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/637,759  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: RPMS 101  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 873-8794  
TELEFAX: (404) 873-8795  
INFORMATION FOR SEQ ID NO: 123:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 56 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
US-09-201-945-123

Query Match 11.3%; Score 50; DB 4; Length 56;  
Best Local Similarity 34.4%; Pred. No. 22;  
Matches 11; Conservative 4; Mismatches 13; Indels 4; Gaps 1;

Oy 41 CHRVSPNSTLKGHVRLCKPKLEPERLM 72

Db 14 CORVERTGOSNLSRYHNR-----CSLNSDVRW 41

RESULT 9  
US-08-726-306A-174

Sequence 174, Application US/08726306A  
Patent No. 5958684

GENERAL INFORMATION:

APPLICANT: van Leeuwen, Frederik Willem

APPLICANT: Burbach, Johannes Peter Henri

APPLICANT: Grosveld, Franklin G.

TITLE OF INVENTION: DIAGNOSIS METHOD AND REAGENTS

NUMBER OF SEQUENCES: 189

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner & Witcoff, Ltd.

STREET: 1 Financial Center

CITY: Boston

STATE: MA

COUNTRY: US

ZIP: 02111

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Wordperfect 6.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/726,306A

FILING DATE: 02-Oct-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 95/20080.4

FILING DATE: 02-Oct-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/009,832

FILING DATE: 01-Jan-1996

ATTORNEY/AGENT INFORMATION:

NAME: Williams, Ph.D., Kathleen M.

REGISTRATION NUMBER: 34,380

REFERENCE/DOCKET NUMBER: 96, 048-A (3255/00784)

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 345-9100

TELEFAX: (617) 345-9111

INFORMATION FOR SEQ ID NO: 174:

SEQUENCE CHARACTERISTICS:

LENGTH: 77 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: unknown

MOLECULE TYPE: peptide

US-08-726-306A-174

Query Match 11.3%; Score 50; DB 2; Length 77;  
Best Local Similarity 24.6%; Pred. No. 32;  
Matches 16; Conservative 6; Mismatches 13; Indels 30; Gaps 3;

Db 21 RRRRRRAATTCRRRRRCRSAAATCGATTRKRMASCSARSHAPAPRR-----RRCRP 72

QY 25 KRRP-AKAMSGRTRLCCH-----RVSPNSTNLKGHHVRLCKP 62

QY 63 CKLEP 67

Db 73 RRATP 77

RESULT 10  
US-08-443-568B-10

Sequence 10, Application US/08443568B

Patent No. 5759807

GENERAL INFORMATION:

APPLICANT: Breese, Tim

APPLICANT: Hayenga, Kirk

APPLICANT: Rindersknecht, Ernst

APPLICANT: Vandlen, Richard

APPLICANT: Daniel, Yansura  
TITLE OF INVENTION: Process for Producing Relaxin  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Penile & Edmonds LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/443,568B

FILING DATE: 22-MAY-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/080,354

FILING DATE: 21-JUNE-1993

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Abrams, Samuel B.

REGISTRATION NUMBER: 30,605

REFERENCE/DOCKET NUMBER: 7842-037

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-493-4935

TELEFAX: 650-493-5556

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 77 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-443-568B-10

Query Match 11.2%; Score 49.5; DB 1; Length 77;  
Best Local Similarity 37.9%; Pred. No. 37;  
Matches 11; Conservative 4; Mismatches 7; Indels 7; Gaps 1;

Db 21 STEGRRAPAKAMSGRTRLCCH-----CCH 42

QY 37 STWSKRKPTGYSGKRRKRLXSALANKCCH 65

RESULT 11  
PCT-US94-06997-10

Sequence 10, Application PCT/US9406997

GENERAL INFORMATION:

APPLICANT: Genentech, Inc.

APPLICANT: 460 Point San Bruno Boulevard

APPLICANT: South San Francisco, California

TITLE OF INVENTION: Process for Producing Relaxin

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESSEE: Penile & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/06997

FILING DATE: 20-JUN-1994

CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Halluin, Albert P.  
REGISTRATION NUMBER: 25,227  
REFERENCE/DOCKET NUMBER: 7842-025-228  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-854-3660  
TELEFAX: 415-854-3694  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 77 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US94-06997-10

Query Match 11.2%; Score 49.5; DB 5; Length 77;  
Best Local Similarity 37.9%; Pred. No. 37;  
Matches 11; Conservative 4; Mismatches 7; Indels 7; Gaps 1;

OY 21 STEGRRPAKAMSGRRRL-----CCH 42  
DB 37 STWSKRKPTGSGKKRQLYSLANKCCH 65

RESULT 12  
5256770-8  
Patent No. 5256770  
APPLICANT: GLASER, CHARLES B.; MORSE, MICHAEL J.; LIGHT, DAVID R.  
TITLE OF INVENTION: OXIDATION RESISTANT THROMBOMODULIN ANALOGS  
NUMBER OF SEQUENCES: 48  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/506,325  
FILING DATE: 09-APR-1990  
SEQ ID NO: 8  
LENGTH: 36

Query Match 10.9%; Score 48; DB 6; Length 36;  
Best Local Similarity 34.5%; Pred. No. 23;  
Matches 10; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

OY 9 LCCVLLCGAVFVSPSEIHARFRGARS 37  
DB 8 LCCVLLCGAVFVSPSEIHARFRGARS 36

RESULT 13  
5244676-2  
Patent No. 5244676  
APPLICANT: BELL, LESLIE D.; MAYER, ERNEST J.; PALMTER, MARK O.; TOLUNAY, H. ESER; WARREN, THOMAS G.; WUN, TZE-CHEIN  
TITLE OF INVENTION: MODIFIED TISSUE PLASMINOGEN ACTIVATOR WITH MODIFIED GLYCOSYLATION SITE  
NUMBER OF SEQUENCES: 9  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/203,047  
FILING DATE: 06-JUN-1988  
SEQ ID NO: 2  
LENGTH: 40

Query Match 10.9%; Score 48; DB 6; Length 40;  
Best Local Similarity 34.5%; Pred. No. 26;  
Matches 10; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

OY 9 LCCVLLCGAVFVSPSEIHARFRGARS 37  
DB 8 LCCVLLCGAVFVSPSEIHARFRGARS 36

RESULT 14  
US-08-743-975-9  
Sequence 9, Application US/08743975  
Patent No. 6057434  
GENERAL INFORMATION:  
APPLICANT: Jian Ni, Guo-Liang Yu and Reiner Gentz  
TITLE OF INVENTION: Mammary Transforming Protein  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
ADDRESSEE: CECCHI, STEWART & OLSTEIN  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07068

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/743,975  
FILING DATE: 01 NOVEMBER 1996  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/006,187  
FILING DATE: 02 NOVEMBER 1995  
ATTORNEY/AGENT INFORMATION:  
NAME: MULINS, J.G.  
REGISTRATION NUMBER: 33,073  
REFERENCE/DOCKET NUMBER: 325800-507 (PF212)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 61 AMINO ACIDS  
TYPE: AMINO ACID  
STRANDEDNESS:  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
US-08-743-975-9

Query Match 10.9%; Score 48; DB 3; Length 61;  
Best Local Similarity 28.3%; Pred. No. 43;  
Matches 13; Conservative 7; Mismatches 16; Indels 10; Gaps 2;

OY 17 FSTSTEGRRPAKAMSGR-----RRLCCHRVSPSPSTLKGHH 56  
DB 13 FVFSINILSRPERWEGMPQSGSGRAKL----LQSPNRKHISTNH 54

RESULT 15  
US-07-689-693B-11  
Sequence 11, Application US/07689693B  
Patent No. 5231011  
GENERAL INFORMATION:  
APPLICANT: David Hillyard  
APPLICANT: Baldomero M. Olivera  
TITLE OF INVENTION: Segregated Folding Determinants  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Thorpe, No. 5231011th & Western  
STREET: 9035 South 700 East, Suite 200  
CITY: Sandy  
STATE: Utah  
COUNTRY: USA  
ZIP: 84070  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 720 KB storage  
COMPUTER: Compaq LTE/286

OPERATING SYSTEM: DOS 4.01  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/689,693B  
FILING DATE: 19910418  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: none  
FILING DATE: na  
ATTORNEY/AGENT INFORMATION:  
NAME: Western, M. Wayne  
REGISTRATION NUMBER: 22,788  
REFERENCE/DOCKET NUMBER: 9925  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (801) 566-6633  
TELEFAX: (801) 566-0750  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 68 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: Prepropeptide sequence for three loop  
NAME/KEY: conotoxin from Circler conotoxin from Conus  
IDENTIFICATION METHOD: Libraries were created  
IDENTIFICATION METHOD: using oligo-dt primed pUC13 vector  
US-07-689-693B-11

Query Match 10.7%; Score 47.5; DB 1; Length 68;  
Best Local Similarity 27.4%; Pred. No. 57;  
Matches 20; Conservative 10; Mismatches 18; Indels 25; Gaps 5;  
QY 6 LSSLCTLLCFSEFST--EGKR--RPAKAWGR-----RTRLCHRVSPNST 50  
DB 4 LGAALLTCLLFSLTAVPLDGDQADQPAORLDRIPTEDPLPLDPKRC-----PPVAC 59  
QY 51 NLKGNHVRCKPC 63  
DB 60 NMG-----CKPC 66

Search completed: May 8, 2003, 21:48:40  
Job time : 19 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 8, 2003, 21:48:13 ; Search time 1349 Seconds

(without alignments)  
972.450 Million cell updates/sec

Title: US-09-599-087b-5

Perfect score: 442  
Sequence: 1 MRLVLSLLCILLCSIF.....PCKLEPEPLWVPGALPOV 81

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 227956

Minimum DB seq length: 25  
Maximum DB seq length: 80

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xip  
-O=/cg21/USPTO.spool/US09599087/runit.07052003.113321.7356/app.query.fasta.1.263  
-DB=EST -OPMT=fastap -SUFFIX=first -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdt -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=25 -MAXLEN=80  
-USBR=US09599087.cgcg.1.1525.0runit.07052003.113321.7356 -NCPU=6 -ICPU=3  
-NO\_XIPY -NO\_MAP -LARGOQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_esttom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_plo:\*  
21: em\_gss\_vit:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_mam:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_tod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	50.5	11.4	60	17	A2918344	A2918344 100600480
C 2	50.5	11.4	69	17	BH639950	BH639950 100803280
C 3	49	11.1	72	14	BQ394194	BQ394194 NISC-ng07
C 4	49	11.1	76	17	A2918366	A2918366 100600480
C 5	49	11.1	79	9	A1877052	A1877052 uc56b11.r
C 6	48	10.9	80	10	AV676540	AV676540 AV676540
C 7	47	10.6	67	17	A2605110	A2605110 1M0426115
C 8	47	10.6	73	10	AW600148	AW600148 SWLCAK10
C 9	47	10.6	74	9	AA739046	AA739046 vv66d04.r
C 10	47	10.6	74	13	BM283913	BM283913 K135804.y
C 11	46.5	10.5	79	9	A1468972	A1468972 t143h02.x
C 12	46	10.4	50	9	AU105245	AU105245 AU105245
C 13	46	10.4	70	14	BQ092063	BQ092063 t36f08.x
C 14	46	10.4	70	14	BQ092271	BQ092271 t239e09.x
C 15	46	10.4	73	14	H22549	H22549 yne69h12.r1
C 16	46	10.4	77	9	AA049559	AA049559 m35b12.r
C 17	45.5	10.3	77	14	BQ666094	BQ666094 pb35910.y
C 18	45.5	10.3	77	14	BQ811778	BQ811778 1030025E1
C 19	45	10.2	67	13	BM518515	BM518515 K193f08.y
C 20	45	10.2	70	17	A2992067	A2992067 2M0276001
C 21	45	10.2	72	14	BQ757145	BQ757145 EBem10_SQ
C 22	45	10.2	72	17	A2493456	A2493456 1M0328G07
C 23	45	10.2	72	17	A2921678	A2921678 1006031C0
C 24	45	10.2	73	9	AA790708	AA790708 vv18f08.r
C 25	45	10.2	79	9	AA496761	AA496761 vn32d12.x
C 26	45	10.2	80	9	AU257386	AU257386 AU257386
C 27	44	10.0	74	13	BM097862	BM097862 EBem04_SQ
C 28	44	10.0	76	9	AU244358	AU244358 AU244358
C 29	44	10.0	78	6	BQ241316	BQ241316 TAE05006F
C 30	43.5	9.8	66	17	BH620911	BH620911 100709960
C 31	43.5	9.8	67	17	AZ808082	AZ808082 2M0071124
C 32	43.5	9.8	72	14	R33083	R33083 yh77c06.s1
C 33	43	9.7	40	9	A1589916	A1589916 tmb1h06.x
C 34	43	9.7	48	17	TA176E12P	TA176E12P T.bruce1
C 35	43	9.7	51	10	AA9232755	AA9232755 f130a01.x
C 36	43	9.7	58	9	AA948249	AA948249 cq33908.s
C 37	43	9.7	58	17	AZ835896	AZ835896 2M0130C22
C 38	43	9.7	62	17	FR0012583	FR0012583 F.tubrl1pe
C 39	43	9.7	64	10	AM626440	AM626440 SMOVAFCAP
C 40	43	9.7	64	10	AM626725	AM626725 SMOVAFCAP
C 41	43	9.7	64	12	BF118532	BF118532 SMOVL3CAN
C 42	43	9.7	70	9	AU258504	AU258504 AU258504
C 43	43	9.7	72	10	AV911900	AV911900 AV911900
C 44	43	9.7	75	9	AL643544	AL643544 AL643544
C 45	43	9.7	75	13	B1753134	B1753134 603025986

## ALIGNMENTS

RESULT 1  
LOCUS A2918344 60 bp DNA linear GSS 17-DEC-2001  
DEFINITION 1006004802.x1 1006 - Rescuemu Grid G Zea mays genomic. DNA  
ACCESSION A2918344  
VERSION A2918344  
KEYWORDS  
SOURCE GSS.  
ORGANISM Zea mays.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1 (bases 1 to 60)  
AUTHORS Walbot,V.

TITLE Maize genomic sequences found using engineered Rescemu transposon  
JOURNAL Unpublished (2001)  
COMMENT Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227  
Fax: 650 725 8221

FEATURES  
source  
Email: walbot@stanford.edu  
Possible ligation site of ends cut by 2 different endonucleases.  
Reverse complemented post-ligation sequence from source sequence.  
Plate: 1006004 row: 36  
Class: transposon-tagged.  
Location/Qualifiers  
1..60  
/organism="Zea mays"  
/cultivar="mixed background W23/A188/B73"  
/db\_xref="taxon:4577"  
/clone\_id="1006 - Rescemu Grid G"  
/tissue\_type="leaf"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: leaf; Vector: Rescemu (engineered from pBluescript backbone); Site\_1: BamHI; Site\_2: BglII; Rescemu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on Rescemu, go to the web site 'www.zmdb.iastate.edu' and follow the links for 'Rescemu'. Grid G was grown at Stanford in 2000. DNA was extracted from leaf punches, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."

BASE COUNT 5 a 17 c 26 g 12 t  
ORIGIN

## Alignment Scores:

Pred. No.: 9.11e+03 Length: 60  
Score: 50.50 Matches: 10  
Percent Similarity: 78.57% Conservative: 1  
Best Local Similarity: 71.43% Mismatches: 2  
Query Match: 11.43% Indels: 1  
DB: 17 Gaps: 1

US-09-599-087b-5 (1-81) x AZ918344 (1-60)

OY 34 G1YATGATGTATATGLeucCysGshisAlrYAlProSerPro 47

DB 43 GGCCACCGCACACACGCGTCATCTGTCACCGG---CCCACGCCA 5

## RESULT 2

BH639950/c

LOCUS BH639950 69 bp DNA linear GSS 14-FEB-2002  
DEFINITION 1008052H07.2EL.y1 1008 - Rescemu Grid I Zea mays genomic, DNA  
sequence.

ACCESSION BH639950

VERSION BH639950.1 GI:18665756

KEYWORDS GSS.

SOURCE Zea mays.

ORGANISM Zea mays

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 69)

AUTHORS Walbot,V.

TITLE Maize genomic sequences found using engineered Rescemu transposon

JOURNAL Unpublished (2001)

COMMENT Contact: Walbot V  
Department of Biological Sciences  
Stanford University  
855 California Ave, Palo Alto, CA 94304, USA  
Tel: 650 723 2227

FEATURES  
source  
Fax: 650 725 8221  
Email: walbot@stanford.edu  
Possible ligation site of ends cut by 2 different endonucleases.  
Reverse complemented post-ligation sequence from source sequence.  
Plate: 1008032 row: 22  
Class: transposon-tagged.  
Location/Qualifiers  
1..69  
/organism="Zea mays"  
/cultivar="mixed background W23/A188/B73"  
/db\_xref="taxon:4577"  
/clone\_id="1008 - Rescemu Grid I"  
/tissue\_type="leaf"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: leaf; Vector: Rescemu (engineered from pBluescript backbone); Site\_1: BamHI; Site\_2: BglII; Rescemu is a 4.9 kb, modified maize Mu transposon designed to allow plasmid rescue from total genomic DNA. Mu elements insert preferentially into transcription units. For more information on Rescemu, go to the web site 'www.zmdb.iastate.edu' and follow the links for 'Rescemu'. Grid I was grown at Berkeley in 2001. DNA was extracted from leaf punches, double digested using BamHI and BglII, and ligated to form circular plasmids. DH10B cells were transformed and then screened on LB plates with ampicillin."

BASE COUNT 7 a 30 c 26 g 6 t  
ORIGIN

## Alignment Scores:

Pred. No.: 1.11e+04 Length: 69  
Score: 50.50 Matches: 10  
Percent Similarity: 66.67% Conservative: 2  
Best Local Similarity: 55.56% Mismatches: 5  
Query Match: 11.43% Indels: 1  
DB: 17 Gaps: 1

US-09-599-087b-5 (1-81) x BH639950 (1-69)

OY 25 LYSATGATGPTGAlaAlaAlaTTPSerGlyATGATGTATGLeucCysGshis 42

DB 62 CGCGGAGGCGGCGCGCGCGGTCGCGC---CGGCTGACCCGAGTTCCTCCAC 12

## RESULT 3

BQ394194/c

LOCUS BQ394194 72 bp mRNA linear EST 22-MAY-2002  
DEFINITION NISC.ng07g11.x1 NICHD XGC Emb6 Silurana tropicalis cDNA clone  
IMAGE:5382740 3', mRNA sequence.

ACCESSION BQ394194

VERSION BQ394194.1 GI:21061881

KEYWORDS EST.

SOURCE western clawed frog.

ORGANISM Silurana tropicalis

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;  
Xenopodinae; Silurana.

1 (bases 1 to 72)

AUTHORS NIH-XGC

TITLE http://image.llnl.gov/image/html/xenopuslib.info.shtml.

JOURNAL National Institute of Child Health and Human Development, National

COMMENT Cancer Institute, Xenopus Gene Collection  
Unpublished (2002)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaps-remail.oh.gov

CDNA Library Preparation by: The I.M.A.G.E. Consortium/LLNL

DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC)

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

info@image.llnl.gov

Plate: LLM11974 row: N column: 21

Seq primer: -21M13 forward primer (ABI).



```

FEATURES
  source
    location/Qualifiers
      1. 72
        /organism="Silurana tropicalis"
        /db_xref="taxon:8364"
        /clone_lib="IMAGE:5382740"
        /dev_stage="NICHD XGC Emb6"
        /tissue_type="neurala"
        /dev_stage="embryo, stages 14-19"
        /lab_host="DH10B (phage-resistant)"
        /note="Vector: pCMV-SPORT6.ccdB; Site_1: NotI; Site_2:
        EcoRV; Cloned unidirectionally. Primer: Oligo dT. Average
        insert size 2.1 kb. Constructed by Invitrogen. Note: This
        is a Xenopus Gene Collection (XGC) library."

BASE COUNT
  13 a 14 c 5 g 40 t

ORIGIN
  Alignment Scores:
    Pred. No.: 1,7e+04 Length: 72
    Score: 49.00 Matches: 9
    Percent Similarity: 65.22% Conservative: 6
    Best Local Similarity: 39.13% Mismatches: 8
    Query Match: 11.09% Indels: 0
    DB: 14 Gaps: 0

US-09-599-087b-5 (1-81) x BQ394194 (1-72)

QY 17 Phseerllephserthrguglylsargargproalalysalatpserglyargarg 36
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 69 TTTAGTGTGTTTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 10
    :::|||||:::

QY 37 Thargleu 39
    :::|||||:::
Db 9 TCTAGAGTA 1

RESULT 4
  A2918366 76 bp DNA linear GSS 17-DEC-2001
  LOCUS 1006004B05.2EL_X3 1006 - Rescemu Grid G Zea mays genomic, DNA
  DEFINITION
  sequence.
  ACCESSION A2918366
  VERSION A2918366.1 GI:13387650
  KEYWORDS GSS.
  SOURCE Zea mays.
  ORGANISM Zea mays.
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
  clade; Panicoidae; Andropogoneae; Zea.
  1 (bases 1 to 76)
  Walbot, V.
  Maize genomic sequences found using engineered Rescemu transposon
  Unpublished (2001)
  Contact: Walbot V
  Department of Biological Sciences
  Stanford University
  855 California Ave, Palo Alto, CA 94304, USA
  Tel: 650 723 2227
  Fax: 650 723 8221
  Email: walbot@stanford.edu
  Possible ligation site of ends cut by 2 different endonucleases.
  Reverse complemented post-ligation sequence from source sequence.
  Plate: 1006004 row: 36
  Class: transposon-tagged.
  Location/Qualifiers
    1. 76
      /organism="Zea mays"
      /cultivar="mixed background W23/A188/B73"
      /db_xref="taxon:4577"
      /clone_lib="11006 - Rescemu Grid G"
      /tissue_type="leaf"
      /dev_stage="adult"
      /lab_host="DH10B"
      /note="Organ: leaf; Vector: Rescemu (engineered from
      pBluescript backbone); Site_1: BamHI; Site_2: BglII;"

```

```

Rescemu is a 4.9 kb, modified maize Mu transposon
designed to allow plasmid rescue from total genomic DNA.
Mu elements insert preferentially into transcription
units. For more information on Rescemu, go to the web
site 'www.zmdb.lastate.edu' and follow the links for
'Rescemu.' Grid G was grown at Stanford in 2000. DNA was
extracted from leaf punches, double digested using BamHI
and BglII, and ligated to form circular plasmids. DH10B
cells were transformed and then screened on LB plates with
ampicillin."

BASE COUNT
  15 a 24 c 22 g 15 t

ORIGIN
  Alignment Scores:
    Pred. No.: 1,84e+04 Length: 76
    Score: 49.00 Matches: 10
    Percent Similarity: 57.89% Conservative: 1
    Best Local Similarity: 52.63% Mismatches: 2
    Query Match: 11.09% Indels: 6
    DB: 17 Gaps: 1

US-09-599-087b-5 (1-81) x A2918366 (1-76)

QY 42 HisarValproserProasnerThrasnleutysglyHisValargleucys 60
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 38 CACTCAATGCATCGCCA-----CCCAACACGTCAGATTATGT 76
    :::|||||:::

RESULT 5
  A1877052 79 bp mRNA linear EST 21-JUL-1999
  LOCUS uc56b11.r1 Soares_thymus_2NBWT Mus musculus cDNA clone
  DEFINITION IMAGE:1429629 5' similar to TR:035500 O35500 KINSEIN LIGHT CHAIN
  ISOFORM D.1, mRNA sequence.
  ACCESSION A1877052
  VERSION A1877052.1 GI:5551101
  KEYWORDS EST.
  SOURCE house mouse.
  ORGANISM Mus musculus
  Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 79)
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP)
  Tumor Gene Index
  Unpublished (1997)
  Contact: Robert Strausberg, Ph.D.
  Email: cgaaps-r@mail.nih.gov
  This clone is available royalty-free through LNL; contact the
  IMAGE Consortium (info@image.llnl.gov) for further information.
  MGI:913697
  Trace considered overall poor quality
  Possible reversed clone; similarity on wrong strand
  Seq primer: -28m13 rev2 BT from Amersham
  High quality sequence stop: 1.
  Location/Qualifiers
    1. 79
      /organism="Mus musculus"
      /strain="C57BL/6J"
      /db_xref="taxon:10090"
      /clone_lib="IMAGE:1429629"
      /clone_lib="Soares_thymus_2NBWT"
      /sex="male"
      /tissue_type="thymus"
      /dev_stage="4 weeks"
      /lab_host="DH10B"
      /note="Vector: pT73D-Pac (Pharmacia) with a modified
      polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
      was primed with a Not I - oligo(dN) primer (5'
      TGTATCCATCTAAGTGGAGGCGCGCTTTTCTTTTCTTTTCTTTTCTTTT
      3'); double-stranded cDNA was ligated to Eco RI adaptors
      (Pharmacia), digested with Not I and cloned into the Not I
      and Eco RI sites of the modified pT73 vector. RNA
      provided by Dr. Bertrand Jordan. Library went through two

```

## RESULTS

MS-09-599-0

87B-5 (1-81) x 27605110





```

FEATURES          a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).
SOURCE            Location/Qualifiers
                  1..50
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /clone="COL05671"
                   /clone_1lb="Sugano Homo sapiens cDNA library"
                   /note="Differential display comparison of untreated and
                        dimethylfumarate treated 0937 cells"
BASE COUNT       2 a 19 c 18 g 11 t
ORIGIN
Alignment Scores:
Pred. NO.:       2.08e+04
Score:           46.00
Percent Similarity: 73.33%
Best local Similarity: 60.00%
Query Match:     10.41%
DB:              9
                Gaps: 0
US-09-599-087B-5 (1-81) x AU105245 (1-50)
Oy 32 TrpSerGlyArgArgThrArgIleucyScySHsArgValProSer 46
Db 5 TTTTCCGCCCGCGCTGCTGCGCTGCTGCGCGCGCTGCCAGT 49
RESULT 13
LOCUS            B0092063/c
DEFINITION       70 bp mRNA linear EST 08-APR-2002
ACCESSION        B0092063
VERSION          B0092063.1
KEYWORDS         GI:20072973
SOURCE            zebrafish.
ORGANISM         Danio rerio
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                  Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
                  Cyprinidae; Danio.
REFERENCE         1 (bases 1 to 70)
AUTHORS          Clark,M., Johnson,S.L., Lebrach,H., Lee,R., Li,F., Marra,M., Eddy
                  ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
                  ,K., Stepien,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
                  Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurt,R., Ritter,E.,
                  Kohn,S., Shih,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
                  and Wilson,R.
TITLE            Mashu zebrafish EST project 1998
JOURNAL           Unpublished (1998)
COMMENT          Contact: Stephen L. Johnson
                  Washington University School of Medicine
                  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
                  Tel: 314 286 1800
                  Fax: 314 286 1810
                  Email: zbrnfish@wustl.edu
                  CDNA Library Preparation: Raymond Lee. CDNA Library Arrayed by:
                  Matthew Clark. DNA Sequencing by: Washington University Genome
                  Sequencing Center Clone distribution: Genome Systems, St. Louis,
                  Missouri (web address: www.genomesystems.com) (email contact:
                  info@genomesystems.com) and Research Genetics, Huntsville, Alabama
                  (web address: www.resgen.com) (email contact: info@resgen.com) and
                  ResourcenzentrumImprimarDatenbank, Berlin, Germany (web address:
                  www.rzpd.de)
                  Seq primer: T7 from GIBCO.
FEATURES          Location/Qualifiers
SOURCE            1..70
                  /organism="Danio rerio"
                  /db_xref="taxon:7955"
                  /clone="5905287"
                  /clone_1lb="zebrafish fin day1 regeneration"
                  /sex="mixed male and female"
                  /tissue_type="1 day fin regenerates"
                  /lab_host="E. coli XL0LR"
                  /note="Vector: pBR-CMV; Site_1: EcoRI; Site_2: XhoI; 1st

```

strand cDNA primed with (GA)10ACTGCTCCAG(T)18, followed by second strand synthesis and ligated to 5' adapter (5' )-atttcgcacgag-3', 3'-gccgtgtc-5'. cDNA was cloned. directionally (EcoRI/XhoI) into Stratagene Zap express lambda phage arms. Mass Invivo excision done to obtain inserts in pBK-CMV phagemid."

BASE COUNT 5 a 19 c 11 g 35 t

ORIGIN

Alignment Scores:

Precl. No.: 70

Score: 46.00

Percent Similarity: 66.67%

Best Local Similarity: 53.33%

Query Match: 10.41%

DB: 14

US-09-599-087B-5 (1-81) x B0092063 (1-70)

Oy 24 G1yLSArGArGPrOAlALySAlArTpSeRGLyArGArGThnArG 38

Db 63 GGAAGAGAGAGAGCCACAAACACAGTGTCTTCAAGACAGACAGA 19

RESULT 14

B0092271/c

LOCUS

DEFINITION

Accession

Version

Keywords

Source

Organism

Danio rerio

Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 70)

Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theisling,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shn,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.

Washu Zebrafish EST Project 1998

Unpublished (1998)

Contact: Stephen L. Johnson

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: zbratfish@watson.wustl.edu

CDNA Library Preparation: Raymond Lee, CDNA Library Arrayed by: Matthew Clark. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: Genome Systems, St. Louis, Missouri (web address: www.genomesystems.com) (email contact: info@genomesystems.com) and Research Genetics, Huntsville, Alabama (web address: www.resgen.com) (email contact: info@resgen.com) and ResourcenzentrumPrimateDatenbank, Berlin, Germany (web address: www.rpdd.de)

Seq primer: T7 from Gibco.

Location/Qualifiers

1..70

/organism="Danio rerio"

/db\_xref="taxon:7955"

/clone="5905624"

/clone\_lib="zebrafish fin day1 regeneration"

/sex="mixed male and female"

/tissue\_type="1-day fin regenerates"

/lab\_host="E. coli XL0LR"

/note="Vector: pBK-CMV; Site\_1: EcoRI; Site\_2: XhoI; 1st strand cDNA primed with (GA)10ACTGCTCCAG(T)18, followed by second strand synthesis and ligated to 5' adapter (5' )-atttcgcacgag-3', 3'-gccgtgtc-5'. cDNA was cloned

